Estimación de parámetros para la toma de decisiones en el proceso de selección de asignaturas en el programa de Ingeniería Civil de la Pontificia Universidad Javeriana

Ricardo Fernando Otero Caicedo

Director
Ing. Juan Pablo Caballero Villalobos
Departamento de Ingeniería Industrial

Pontificia Universidad Javeriana
Ingeniería Industrial
Mayo 20 de 2013

Tabla de contenido

1 Glosario

Base de datos académica depurada: Base de datos a la cual se le ha aplicado la metodología de pre-proceso y depuración indicada en la sección 9.4.

Coeficiente de asimetría: Indicador que estima el grado de simetría de un conjunto de datos con respecto a su media. Permite identificar si la frecuencia de los datos está concentrada en la media o en un valor menor o mayor a ella.

Coeficiente de correlación de Pearson: Es un indicador que mide la relación lineal entre dos variables cuantitativas. Su rango va de $[-1,1]$, un valor positivo indica una relación lineal directamente proporcional y un valor negativo indica una relación lineal inversamente proporcional. Entre más alejado esté el indicador del 0 se presenta una mayor correlación.

Coeficiente de kurtosis: Es un indicador de forma de un conjunto de datos, un valor entre 1.5 y 3.0 del coeficiente de kurtosis indica una concentración de datos alrededor de la media con una gran frecuencia de datos extremos.

CP: Asignaturas complementarias.

CSV: Archivo plano donde cada dato se separa por coma (,) ó punto y coma (;).

ELE: Asignaturas electivas.

ENF: Asignaturas de énfasis.

FDP: Función de probabilidad.

GLPK: Es un paquete de software dirigido a la solución de problemas de programación lineal, entera y mixta de gran escala.

GNU: Movimiento y comunidad de conocimiento libre, con el objetivo del desarrollo colaborativo de software mediante el uso de licencias *open source*.

GA: Algoritmo genético.

Grafo: Es un gráfico representado por un conjunto de objetos conocidos como nodos los cuales están unidos por arcos que pueden indicar algún tipo de relaciones entre ellos.

GUI: Interfaz gráfica de usuario.

IDM: Número de identificación de una asignatura.

KN: *Knapsack problem*, problema combinatorio relacionado con la selección de un conjunto de objetos que tienen un aporte de utilidad y un costo asociado.

***Kolmogorov –smirnov*:** Estadístico de prueba de la bondad de ajuste de un conjunto de datos a una distribución de probabilidad en particular.

NFF: Asignaturas núcleo fundamental.

Percentil: El percentil *i* se refiere al valor de una variable bajo el cual se encuentra un i-ésimo porcentaje del total de observaciones.

Relajación lineal: Aproximación de un modelo de programación matemática que intenta simular las mismas condiciones del problema original pero convirtiendo sus restricciones y/o objetivo en funciones lineales con el fin de facilitar su análisis y solución.

Significancia: Un resultado se asume que es estadísticamente significativo cuando su valor de significancia también conocido como valor P o mínimo alfa es menor a un valor determinado por el investigador que usualmente toma valores desde el 1% hasta el 10%. Si un valor es estadísticamente significativo indica que no es probable que este resultado haya sido resultado del azar.

SIU: Sistema de información universitaria.

2 Introducción

La flexibilización de los sistemas de educación superior ha contribuido en la interacción transversal de los componentes centrales de cada programa académico con diferentes áreas del conocimiento, desarrollando así capacidades globales y permitiendo conexión y sinergias con profesionales de otras disciplinas [1]. El empoderamiento hacia los estudiantes en la estructuración de su propio plan de estudios ha permitido satisfacer los objetivos enfocados a captar conocimiento, paralelo a una educación integral que asegure espacios de formación investigativa y creativa.

Actualmente el esquema de educación en la Pontificia Universidad Javeriana está basado en el sistema de créditos académicos. Según la Vicerrectoría Académica, un crédito corresponde a "la unidad que mide la actividad del estudiante y que pondera equilibradamente los siguientes criterios: Número total de horas de trabajo académico, tipo de trabajo asistido, grado de dificultad de la asignatura y su importancia dentro del plan de estudios" [2]. Dentro del sistema de créditos académicos se permite la selección flexible de las asignaturas del plan de estudios, restringido únicamente por el número total de créditos por matrícula y las condiciones específicas de cada asignatura.

Desde la implementación formal del sistema de créditos en la Pontificia Universidad Javeriana en el año 2002, se han definido claramente los siguientes aspectos: Criterios de definición de matrícula académica completa y media, identificación de las horas dedicadas por asignatura (en función de su número de créditos) y ponderación de promedios académicos. Para contribuir al proceso educativo centrado en este sistema, actualmente, cada programa brinda apoyos didácticos para guiar a los estudiantes en la selección del currículo académico de cada periodo, que permita un desarrollo cognitivo y satisfaga los requisitos de la Universidad y a su vez las convicciones de aprendizaje propias de cada estudiante.

La conformación del conjunto de asignaturas a inscribir en un periodo requiere que el estudiante establezca las opciones que considere pertinentes y las analice según el direccionamiento provisto por las herramientas de apoyo de cada programa y su experiencia previa. Este es un proceso de decisión complejo debido a que requiere analizar una amplia cantidad de posibles conjuntos de asignaturas guiado por criterios objetivos y metas precisas. Para poder sobrellevar este proceso de decisión, es preciso disponer de información relevante y objetiva que soporte la situación académica particular de cada estudiante y que además, asegure que las diferentes opciones a tener en cuenta contribuyan con el avance en el programa académico y el cumplimiento de los requisitos exigidos por la universidad.

Para que los estudiantes sean beneficiados por la flexibilidad curricular que expresa el sistema de créditos académicos, es necesario que sea asesorado en el proceso de construcción del plan de estudios que va a desarrollar a lo largo de la carrera, conformándolo adecuadamente de acuerdo con sus requerimientos, convicciones y restricciones. En las universidades existen ciertos lineamientos institucionales que apoyan este proceso de decisión para que los estudiantes no cometan errores al momento de crear su plan de estudios, pero, a pesar de que estos lineamientos existen y son coherentes, sería de gran apoyo contar con una metodología formal y objetiva que responda de manera eficaz cada situación en particular.

En cada una de las facultades de la Universidad se registran datos que comprenden diferentes variables que sintetizan los resultados y elecciones producto de las decisiones de los estudiantes en cada periodo académico. Si bien es cierto que por lo general esta información no está dirigida fuera de los procesos administrativos de control y seguimiento en las Universidades [3], es preciso brindar protagonismo a la información contenida en estas bases de datos para que apoyen los procesos académicos internos como la selección de asignaturas periodo a periodo por parte de los estudiantes.

Como caso de estudio, en este proyecto se pretende proponer una metodología que permita extraer información acertada de las bases de datos que contienen los registros académicos para así estimar parámetros pertinentes que brinden información objetiva y estandarizada de cada una de las principales asignaturas del programa académico. De igual manera se propone un modelo matemático que intenta simular el proceso de selección de asignaturas, sujeto a los requisitos impartidos por la Universidad que brinde decisiones acertadas y estandarizadas, basadas en el perfil y situación específica de cada estudiante. Debido a la complejidad del modelo se presenta un método de solución apropiado para este caso basado en programación matemática. Todo esto, utilizando la información específica suministrada por el programa de Ingeniería Civil de la Pontificia Universidad Javeriana sede Bogotá.

Se espera que la implementación de esta metodología de apoyo impacte positivamente en las expectativas personales de los estudiantes, para generar así un mejor aprovechamiento del sistema educativo que se propone en la universidad.

3 Planteamiento del problema

Usualmente el proceso actual de selección de asignaturas que realiza un importante número de estudiantes, se efectúa de manera empírica a través de la interacción y contacto con estudiantes en situaciones similares, quienes estructuran sus reglas de elección a través de las experiencias pasadas producto de la percepción hacia las asignaturas ya cursadas.

Antes de iniciar cada periodo académico, los estudiantes de segundo semestre en adelante deben elegir e inscribir el grupo de asignaturas que consideren más recomendables según sus objetivos, los cuales pueden estar principalmente relacionados con: convicciones de aprendizaje, cumplimiento del plan de estudios, promedio ponderado y carga académica. Este es un proceso de selección complicado en el que además de contemplar una gran cantidad de posibilidades, existe alta subjetividad en la aplicación de reglas empíricas de elección.

Más aún, existen situaciones críticas como estudiantes en periodo de prueba académica, en difícil situación económica, en búsqueda de beneficios especiales como matrícual de honor, becas, inscripción de asignaturas cooterminales, etc.; en las que es fundamental tomar decisiones acertadas sobre la

configuración de asignaturas del periodo académico siguiente con el fin de incrementar la posibilidad de asegurar sus propósitos académicos y personales.

En la actualidad, se pueden identificar los lineamientos institucionales académicos y las figuras de consejeros académicos como herramientas de apoyo en el proceso de decisión curricular efectuado por los estudiantes cada semestre.

Los lineamientos institucionales académicos establecen entre otros, la malla curricular, la cual muestra de manera detallada un ejemplo general de cómo podrían inscribir los estudiantes las asignaturas periodo a periodo; de esta manera, la malla curricular le sirve como guía para que el estudiante ajuste sus decisiones según su plan de carrera futuro, los requisitos de la Universidad y su convicción de aprendizaje.

Por otra parte, la consejería académica es un servicio de atención particular que realiza un profesor de la Universidad perteneciente al programa académico, en el cual se analizan las diferentes opciones que tiene el estudiante y se toma una decisión conjunta. En otras palabras, el consejero es un orientador que guía hacia la creación de alternativas viables, según su juicio y experiencia, en conjunto con las convicciones propias del estudiante.

Es recomendable que los estudiantes combinen las estregias de acompañamiento de la Universidad con sus convicciones de aprendizaje, para lo cual, es necesario validar métodos que permitan la toma decisiones acertadas, tratando el proceso de manera objetiva y basándose en información pertinente y precisa. Además, estos métodos deben simuar el contexto de la carrera de ingeniería civil, entre los cuales encontramos los requisitos de precedencia de asignaturas, número de créditos aprobados, requisitos de idiomas y diferentes tipos de asignaturas.

Actualmente, cada una de las asignaturas puede clasificarse por su obligatoriedad y propósito en diferentes grupos acorde al papel que cumplen dentro del programa académico. El núcleo de formación fundamental (NFF) incluye todas las asignaturas que se consideran indispensables para la formación profesional en las diferentes áreas de la carrera. Las asignaturas de énfasis (ENF) son aquellas que profundizan el conocimiento en un área específica de la carrera que escoge el estudiante entre las opciones implementadas por la carrera de Ingeniería Civil. Las asignaturas de opción complementaria (CP) promueven la apropiación y aplicación de conocimientos de un campo específico, en otra área del conocimiento, lo que le permite al egresado un complemento en su formación disciplinar. Las asignaturas electivas (ELE) de flexibilización contribuyen a la formación integral a través de cursos de formación humanística, deportiva, técnica, cultural entre otros [4].

Para poder estimar el posible efecto de tomar una decisión en particular, es necesario buscar una representación válida que indique la consecuencia de elegir ese conjunto de asignaturas y que sirva de base para compararla con otras posibles alternativas. Actualmente, el indicador a través de la cual se mide el desempeño de los estudiantes es el promedio académico ponderado acumulado, el cual se refiere a la ponderación de las notas de las asignaturas cursadas y el número de créditos académicos de cada una de ellas. Esta variable depende de los resultados obtenidos en cada una de las asignaturas finalizadas hasta el momento, por lo que servirá como una de las medidas de ajuste de cada una de las posibles alternativas. Así mismo, es necesario cuantificar el aporte de cada asignatura en el cumplimiento total del plan de estudios, teniendo en cuenta las necesidades específicas de cursar y aprobar las asignaturas de núcleo de formación fundamental, énfasis, complementarias y electivas.

También es necesario restringir los conjuntos de asignaturas que se pueden elegir, excluyendo aquellos que no cumplan con los requisitos inherentes al problema, particularmente, cada conjunto de opciones deberá estar limitado según el número de créditos a cursar de acuerdo a la modalidad que el estudiante desee matricular para el periodo y cada asignatura deberá cumplir con los requisitos de precedencia, además de los particulares de algunas asignaturas como número mínimo de créditos aprobados, cumplimiento de requisito de lengua extranjera.

Estas características del problema hacen que sea complicado encontrar la mejor alternativa dentro del conjunto total de posibilidades, puesto que, el número total de opciones es considerablemente grande ya que proviene de la operación combinatoria entre todas las asignaturas que el estudiante pudiera inscribir y el cardinal de subconjuntos que es posible conformar sin infringir la restricción asociada al número de créditos y relaciones de dependencia. En particular, el número total de opciones estará dado por:

$$\sum_{i=1}^{N} \binom{N}{i} = 2^N - 1$$

En donde N es el número total de opciones que el estudiante puede inscribir en un periodo determinado.

Una de las maneras más sencillas de representar una solución es escogiendo entre dos alternativas, que en este proyecto representan la elección o rechazo de incluir una asignatura candidata dentro del conjunto de inscripción sugerido. Este tipo de decisiones puede ser formulado dentro de un modelo matemático como una variable binaria, es decir $x \in \{0,1\}$ [4], donde 1 representa la selección de la asignatura dentro del conjunto.

Por lo expuesto anteriormente y una vez teniendo estipulado cuál debe ser la metodología de estimación de parámetros que permita abordar el problema adecuadamente, el interrogante del proceso de selección de asignaturas se podría definir de la manera siguiente: ¿Qué asignaturas se deben elegir de un gran conjunto de posibilidades, que maximicen el valor esperado de la función objetivo, restringido al número de créditos permitido y características de cada asignatura?

Haciendo una analogía al problema presentado anteriormente, a continuación se presenta el problema clásico de programación matemática, específicamente de programación entera, el problema de la mochila o *Knapsack problem (KN)*, el cual se puede ilustrar con el siguiente ejemplo provisto por *Hans Kellerer* [5]: Considere un montañista quien está empacando su mochila para un viaje a la montaña y tiene que decidir cuáles ítems debería llevar en el viaje. Él tiene una gran cantidad de artículos para elegir, los cuales podrían ser útiles en el viaje. Cada uno de esos artículos numerados de $\mathbf{1}$ a $\boldsymbol{n}$ le darán cierta cantidad de beneficio por elegirlos, el cual está expresado como $\boldsymbol{p_j}$. Por supuesto cada uno de los ítems que lleve con él aumentará el peso de la mochila en una cantidad $\boldsymbol{w_j}$**.** Por obvias razones el peso total de los artículos que lleve al viaje no podrá sobrepasar la capacidad $\boldsymbol{c}$ de la mochila.

La situación descrita anteriormente es una representación equivalente al problema de elección de asignaturas. Las restricciones de peso de la mochila representan el número de créditos máximo por semestre, mientras que los beneficios de cada ítem indican el aporte del resultado de la calificación de esa asignatura dentro del promedio total ponderado acumulado. De esta manera este proyecto está enmarcado dentro de los desarrollos de las investigaciones disponibles realizadas hasta el momento en la modelación, implementación y métodos de solución de las diferentes representaciones del KN.

El presente trabajo de grado busca responder las siguientes preguntas de investigación: ¿Cuál debe ser la metodología adecuada de estimación de parámetros que permitan sustentar el proceso de selección de asignaturas de los estudiantes del programa de Ingeniería Civil?, ¿Cuál debe ser la representación del problema de decisión que permita incluir modelos de optimización? Esto enmarcado dentro de los lineamientos institucionales de la Pontificia Universidad Javeriana sede Bogotá y la convicción de aprendizaje de los estudiantes.

4 Antecedentes

4.1 Iniciativas Institucionales

La implementación de currículos flexibles ha obligado a crear espacios y herramientas que permitan beneficiar el desarrollo de la madurez cognitiva y a su vez identificar intereses particulares que favorezcan el despliegue del potencial de los estudiantes enmarcado en un ambiente de formación integral [2]. Estas estrategias están enfocadas principalmente a generar un acompañamiento a los estudiantes que permita guiarlos hacia un camino común entre los requerimientos académicos de la Universidad y las convicciones de aprendizaje propias de cada estudiante.

4.1.1 Estrategias de acompañamiento

A continuación se presentan algunas estrategias recopiladas según la revisión bibliográfica.

4.1.1.1 Consejería académica

Es un acompañamiento que brinda la Universidad para facilitar las decisiones de los procesos de aprendizaje, a través del análisis y estructuración del plan de estudios. Según lo expuesto en la Vicerrectoría Académica [6]: “En la relación profesor-estudiante, es fundamental que el docente aporte junto con su calidad humana y madurez, su competencia académica y experiencia en la participación de los procesos de enseñanza”.

4.1.1.2 Malla curricular

Responde a la necesidad de integrar las asignaturas ordenándolas cronológicamente por semestres según una estructura lógica y metodológica en la que están contenidos cada uno de los componentes del programa educativo, y en la que se ve discriminado cada una de las asignaturas del plan de estudios con su respectivo número de créditos y relaciones de precedencia. También brinda información acerca de los componentes de núcleo de formación fundamental, énfasis, opciones complementarias y electivas interdisciplinares.

4.1.1.3 Decisiones guiadas paso a paso

Guías generales que permiten a los estudiantes analizar cuáles son los criterios que deben tomar en cuenta al momento de decidir qué asignaturas cursar el siguiente semestre [7]. Indican cuáles criterios deben tener los estudiantes para estructurar adecuadamente los cursos necesarios para cumplir con los requisitos académicos y también estructuran el pensamiento para motivar a los estudiantes a identificar sus convicciones de aprendizaje extracurricular.

4.1.1.4 Herramientas de apoyo generales

Grupos de decisión, tutoriales multimedia, seminarios, entre otras herramientas que pueden apoyar y centrar el proceso de decisión de los estudiantes semestre a semestre [8]. En los videos y seminarios se

incluyen a directores de programas académicos quienes comentan sobre las experiencias que han tenido y cómo creen ellos que debe ser el proceso de selección de asignaturas.

4.2 Iniciativas de modelación y métodos de solución del Knapsack Problem

4.2.1 Modelación

A continuación se presentan de manera general las principales variaciones que ha tenido el problema KN y los diferentes métodos de soluciones aplicados según las investigaciones recolectadas en la revisión bibliográfica.

4.2.1.1 Problema de la mochila (KN)

El problema de la mochila (KP) es un caso particular de los problemas binarios, el cual se define formalmente así: Se tiene un conjunto **N** de ítems, cada ítem **j** tiene asociado una función de utilidad $\boldsymbol{p_j}$ y un peso $\boldsymbol{w_j}$. La mochila tiene una capacidad máxima de peso **c**, por lo que el problema consiste en escoger el subconjunto de ítems que maximiza la función de utilidad total [9]. Este tipo de problema es útil para modelar situaciones como: Programación de producción, selección de maquinaria, inversión en el mercado de valores, etc. [10], [4], [5]. Los problemas KN han sido ampliamente estudiados, especialmente en las últimas décadas, atrayendo a teóricos y a desarrolladores. Los intereses en este tipo de problemas radican en su estructura simple, la cual permite explotar las propiedades combinatorias y también cualquier otro tipo de variantes complejas implementadas al problema base [4].

De manera general, los KN pueden expresarse matemáticamente de la siguiente manera:

$$\max \sum_{j=1}^{n} p_j x_j$$

$$S.A. \sum_{j=1}^{n} w_j x_j \leq c$$

$$x_j \in \{0,1\}\ \forall j$$

En donde $x_j = 1$, representa que el ítem j se incluye dentro de la mochila y $x_j = 0$, si no es tenido en cuenta en el subconjunto.

En la revisión bibliográfica realizada se encontraron principalmente algunas variaciones del problema KN que podrían estar relacionadas con las necesidades de este proyecto.

4.2.1.2 KN multi-objetivo (MOKN)

Los problemas de optimización combinatoria con más de una función objetivo son complicados de abordar, dado que muchos grupo de soluciones mejorarán alguna función objetivo pero sacrificando en la mayoría de las veces las demás [11]. Es más, una solución que sea óptima para las demás funciones objetivo podría ser no factible. En particular, el problema KN multi-objetivo con $\boldsymbol{q}$ funciones objetivo puede representarse de la siguiente manera:

$$\max \quad f_k(x) = \sum_{j=1}^{n} p_{jk} x_j\,, \quad k = 1, \dots, q$$

$$S.A. \sum_{j=1}^{n} w_j x_j \leq c$$

$$x_j \in \{0,1\}\ \forall j$$

Este tipo de formulación permite adaptarse de mejor manera a las situaciones de la vida real, en las cuales es necesario mejorar varias situaciones dependientes de los ítems a seleccionar. Aplicaciones reales se han realizado en selección de inversión en sistemas de transporte, reubicación para la conservación biológica, planificación y remediación de estaciones de energía, entre otras [12].

4.2.1.3 KN con restricciones múltiples (MOMCKN)

En varias ocasiones cada subconjunto de ítems puede tener restricciones adicionales al peso total de la mochila. Estas restricciones podrían incluir relaciones entre los ítems, las cuales deben ser formuladas con programación binaria. En particular este problema puede formularse de manera similar al MOKN modificando las restricciones [13].

$$S.A. \sum_{j=1}^{n} w_{ij} x_j \leq c_i$$

$$\sum_{j=1}^{n} f_j(x_j) \leq k_j$$

$$x_j \in \{0,1\}\ \forall j$$

Este tipo de formulaciones sirve para poder representar relaciones adicionales entre los ítems que pudieran ser seleccionados para ser incluidos dentro del subconjunto final.

4.2.1.4 KN no lineal (NLKN)

Las funciones objetivo $f(x)$ y las restricciones $g(x)$ son no lineales, continuas y diferenciables. De manera general el NLKN se puede formular de la siguiente manera:

$$\max \sum_{j=1}^{n} f_j(x_j)$$

$$S.A. \sum_{j=1}^{n} g_j(x_j) \leq c$$

$$x_j \in \{0,1\}\ \forall j$$

Los problemas no lineales tienen una gran cantidad de aplicaciones en diferentes campos como: Selección de portafolio, distribución de recursos, planeación de la producción, redes computacionales, planeación de capacidad en procesos de manufactura, entre otros. [14] [15].

4.2.1.5 KN estocástico

En la mayoría de situaciones de la vida real los problemas son no-determinísticos debido a que no se conoce el valor exacto de algunos de los parámetros en el momento en el que se debe tomar la decisión. Este tipo de modelación es recurrente cuando las restricciones y/o las funciones objetivo tienen

componentes inciertos o estocásticos pero con alguna distribución de probabilidad asociada [16]. La implementación de esta clase de componentes dentro del modelo permite emular el comportamiento real de todas las posibles características involucradas en el modelo que tiene dependencia directa del azar o que están en función de parámetros y variables que tienen definida un comportamiento probabilístico.

4.2.2 Métodos de solución

A continuación se presentarán los métodos de solución que han sido aplicados para resolver las diferentes configuraciones del problema del KN, según la revisión bibliográfica.

4.2.2.1 Ramificación y acotamiento (Branch and Bound)

Los métodos de ramificación y acotamiento están basados en la noción de enumerar de manera inteligente y adecuada todas las posibles soluciones de un problema. Específicamente el método busca construir y probar que una solución es óptima basada en la sucesiva partición inteligente del espacio solución. Ramificación se refiere al proceso de partición y el término acotamiento hace alusión a las cotas que se construyen para probar la optimalidad de una solución sin necesidad de hacer una búsqueda exhaustiva. Es común que los algoritmos de ramificación y acotamiento terminen antes de encontrar el óptimo ya sea por diseño o necesidad, es más común especificar reglas de detención del algoritmo asociadas a la de cercanía a las cotas o al tiempo de ejecución [17].

El algoritmo consiste en una enumeración profunda de una parte de las soluciones al problema, la estrategia de ramificación consiste en seleccionar un ítem $\mathbf{j}$ y generar dos nodos $\mathbf{x_j = 1}, \mathbf{x_j = 0}$. Dependiendo del conocimiento que se tenga del problema se deben crear reglas especiales que permitan discriminar los nodos que seguirán explorando; las cotas superiores son definidas frecuentemente a través de la relajación lineal continua del problema [17]. Existe una gran cantidad de estudios relacionados con técnicas *Branch and Bound*, específicamente en aquellos que crean modelos híbridos con otras herramientas de optimización, entre estos los trabajos realizados por *Galimyanova* [18], *Graham et al.* [19], *Dyer et al.* [20].

4.2.2.2 Programación dinámica

La programación dinámica está asociada a los métodos de ramificación y acotamiento dado que también desarrolla una numeración inteligente del espacio solución de un problema aunque lo hace de una manera diferente. La idea es trabajar desde las últimas decisiones hasta las primeras. Si se necesita establecer una secuencia de $\boldsymbol{n}$ decisiones para resolver un problema de optimización combinatorio, entonces si $\boldsymbol{D_1, D_2, D_3, \dots, D_n}$, es una secuencia óptima, entonces las últimas $\boldsymbol{k}$ decisiones son óptimas igualmente, es decir, el complemento de una secuencia óptima debe ser óptimo igualmente. En particular la programación dinámica divide el problema en pequeñas fases en las cuales se debe tomar una decisión basados en una relación de recurrencia hacia las fases posteriores del problema [17], [21], [22].

Para instancias pequeñas del problema, es posible formular el KN como la integración de varios sub-modelos, los cuales se pueden resolver de manera individual para luego consolidar la solución óptima a través de la integración recursiva de los resultados del sub-problema. Los primeros resultados de la programación dinámica al problema KN fueron presentados por *Martello* [23], en los cuales se muestran los resultados de tiempos de ejecución del modelo. Actualmente la programación dinámica se usa en modelos híbridos que relacionan otras técnicas de solución como meta-heurísticas y ramificación acotamiento [24].

4.2.2.3 Heurísticas

La complejidad de los problemas KN han creado la necesidad de implementar procedimientos específicos para resolver clases de problemas en tiempos razonables. Entre algunos de los algoritmos creados, se destacan los trabajos realizados por *Martello y Torth* [4]*, Graham y Joux* [19]; y las investigaciones implementando heurística *Greedy* [24].

4.2.2.4 Meta-heurísticas

La creación y validación de nuevos modelos meta-heurísticos impulsado por el aumento de la potencia computacional ha incrementado la exploración de estos métodos de solución a problemas complejos, especialmente aquellos de naturaleza combinatoria [25]. Según el estudio de *Jones* llevado a cabo en el año 2002 [26], el 70% de los artículos usa algoritmos genéticos, 24% recocido simulado, y 6% búsqueda tabú. Dentro de la revisión bibliográfica los trabajos de implementación de Meta-heurísticas son [26], [27], [28], [29], [30].

4.2.3 Síntesis KN

A manera de síntesis se presentan a continuación las características de los artículos consultados dentro de la revisión de bibliográfica, con el fin de analizar las particularidades de las investigaciones que pueden estar relacionadas con los objetivos de este proyecto.

Tabla No. 1 Clases de problemas KN y el método de solución aplicado según la revisión bibliográfica.

Artículo	Tipo Problema	Función Objetivo	# Funciones Objetivo	# Restricciones	Método de Solución	Naturaleza de las Variables
[26]	KN Multidimensional	Lineal	Múltiples	Múltiples	Meta-heurística: Algoritmos Genéticos	Determinística
[10]	KN Problema *Online*	Lineal	Única	Única	Heurística: Algoritmo competitivo constante	Orden de llegada estocástico
[22]	KN Multidimensional	Lineal	Única	Múltiples	Programación dinámica con técnicas de reducción	Determinística
[11]	KN Multi-objetivo	Lineal	Múltiple	Única	Esquema de aproximación en tiempo polinomial	Determinística
[12]	KN Multi-objetivo	Lineal	Múltiple	Única	Programación dinámica y reglas de dominancia	Determinística
[27]	KN Multi-objetivo	Lineal	Múltiple	Única	Meta-heurística: Path Relinking	Determinística
[14]	KN No lineal	No Lineal	Única	Única-Múltiple	Heurísticas: Métodos de Búsqueda	Determinística
[28]	KN sin Cotas (Unbounded)	Lineal	Única	Único	Meta-heurísticas: algoritmos genéticos	Determinística
[15]	KN separable no Lineal	No lineal	Único	Único	Heurísticas	Determinística
[20]	KN de opción múltiple	Lineal	Única	Múltiple	Híbrido: Programación dinámica, ramificación y acotamiento	Determinística
[13]	Multidimensional – Multiobjetivo	Lineal	Múltiple	Múltiple	Programación matemática y algoritmos evolucionarios	Determinística

Artículo	Tipo Problema	Función Objetivo	# Funciones Objetivo	# Restricciones	Método de Solución	Naturaleza de las Variables
[18]	KN Simple	Lineal	Única	Única	Programación dinámica y ramificación y acotamiento	Determinística
[24]	KN con pesos binarios	Lineal	Única	Única	Heurística Greedy	Determinística
[19]	Problemas difíciles de KN	Lineal	Única-Múltiple	Única-Múltiple	Algoritmos con métodos exactos	Determinística
[16]	KN estocástico	No Lineal	Única	Única	Heurística	Estocástica
[29]	KN Dos dimensiones	Lineal	Única	Múltiple	Meta-heurística: Recocido Simulado	Determinística
[30]	KN con pesos imprecisos	No lineal	Única	Única	Meta-heurísticas: Algoritmos genéticos	Determinística
[23]	KN simple	Lineal	Única	Única	Algoritmos implementados con Programación dinámica y Ramificación y acotamiento	Determinística
[21]	KN simple	Lineal	Única	Única	Programación Dinámica	Determinística

5 Objetivo general

Proponer una metodología para el análisis de datos del programa de Ingeniería Civil con el fin de estimar parámetros de entrada de métodos de optimización que soporten el proceso de elección de asignaturas principalmente de los estudiantes del programa que estén en prueba académica.

6 Objetivos específicos

- Proponer un modelo matemático para estimar el conjunto de asignaturas a seleccionar por parte de un estudiante del programa.
- Identificar y caracterizar variables involucradas y las relaciones existentes entre ellas.
- Desarrollar un prototipo de aplicación para la selección de asignaturas.
- Proponer una metodología de análisis de datos acorde al problema abordado.

7 Marco teórico

A continuación se presenta una breve introducción a los tópicos en los cuales estará enmarcado el desarrollo de este proyecto:

7.1 Estadística descriptiva

Área de la estadística que estudia la interpretación de las características de una población a través de una muestra mediante el proceso de contextualizar y obtener información de un gran conjunto de datos utilizando técnicas de recolección, limpieza y análisis [31]. Con las herramientas de la estadística descriptiva es posible entonces entender el conjunto de datos en estudio para asignarle cualidades que permitan ampliar su entendimiento y caracterizarlo. Algunas de las herramientas de la estadística descriptiva son:

- **Parámetros de tendencia central:** Son números que expresan una característica en particular de una población. Entre las más reconocidas se encuentran la media aritmética o promedio, desviación estándar, moda, percentiles, rango, kurtosis, asimetría.
- **Histograma:** Muestra la tendencia central de los datos, la dispersión y la forma en cómo estos se distribuyen. El histograma se consigue dividiendo el eje horizontal en intervalos y trazando un rectángulo sobre cada uno de ellos en el cual su área es proporcional al número de observaciones con valores incluidos en ese rango.
- **Gráficos de cajas:** Es una manera visual y útil de representar los datos gráficamente. Éste muestra los valores del percentil no. 5 y no. 95, los cuartiles inferiores y superiores y la mediana. La caja está formada desde los cuartiles inferior y superior, mientras que los bigotes se muestran en el percentil no. 5 y no. 95. Este gráfico permite observar el rango, dispersión y asimetría de los datos.
- **Gráficos de dispersión o puntos:** Permite analizar de manera inmediata la localización o tendencia de los datos y su dispersión. En dos dimensiones se consigue graficando los puntos de las parejas ordenadas (x, y).

7.2 Distribuciones de probabilidad

Si bien los gráficos presentados en el apartado 7.1 anteriormente son útiles para resumir la información de una muestra de datos, es necesario en algunas ocasiones representar las muestras con una distribución de probabilidad para obtener más detalle de su comportamiento. Una distribución de probabilidad es una función que asigna la probabilidad de que una variable aleatoria tome un valor en particular [32]. Existen diversas distribuciones de probabilidad, tanto discretas como continuas, entre ellas: Uniforme discreta, binomial, normal, beta, etc.

- **Pruebas de bondad de ajuste:** Permiten analizar qué tan bien se ajustan los datos a una distribución de probabilidad en particular. Entre los métodos más reconocidos está el estadístico de *Kolmogorov-Smirnov, Shapiro Wilk* y la prueba Chi Cuadrado.

7.3 Inferencia estadística

Es un área de la estadística que comprende los métodos y procedimientos para obtener conclusiones generales para una población a partir del estudio de una muestra. Entre algunas técnicas de inferencia estadística se encuentra: Estimación de parámetros de distribuciones de probabilidad, contrastes de hipótesis, análisis de regresión [31].

7.4 Análisis de regresión lineal

La relación entre una variable que depende de ***k*** variables explicativas caracterizada en un modelo matemático corresponde a un modelo de regresión. Entonces un modelo de regresión lineal con ***k***

predictores describe el hiperplano en el espacio de $\boldsymbol{k}$ dimensiones que representa a la variable dependiente en función lineal y aditiva de sus variables explicativas [32]. El modelo básico de regresión lineal se puede expresar así:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_k x_k$$

En donde $\boldsymbol{\beta_0}$ representa el corte con el eje x cuando el valor de las variables explicativas es 0 y β_k es un parámetro del modelo que indica el cambio en la variable dependiente por cada unidad de la variable explicativa $\boldsymbol{x_k}$ [33].

7.5 Regresión paso a paso (stepwise regression)

Es un procedimiento usado para determinar de manera automatizada las variables explicativas que deberían estar presentes en el modelo de regresión. Este procedimiento permite agregar únicamente las variables necesarias evitando traslapes e imprecisión al modelo final [34]. Este tipo de herramientas permiten disminuir el tiempo de análisis de datos y la selección manual de variables y criterios a tener en cuenta dentro del modelo de regresión. Existen principalmente dos métodos, regresión hacia adelante y hacia atrás. En el primero, el modelo empieza sólo con la constante $\boldsymbol{\beta_0}$ y agrega una a una las variables que indiquen un mejoramiento en el ajuste del modelo, mientras que en el segundo método, se incluyen todas las variables y se van retirando las que no se consideren significativas. Los criterios más usados para agregar o eliminar variables son la prueba F, R cuadrado y el criterio bayesiano [35].

7.6 Programación matemática

Conjunto de teoremas, métodos y técnicas con el objetivo de minimizar o maximizar una función objetivo en presencia de algunas restricciones asociadas. Los problemas de optimización tienen asociado un objetivo modelado matemáticamente al cual se desean encontrar los valores máximos o mínimos de acuerdo a unas restricciones que son impuestas por la misma naturaleza del problema. Según sea la naturaleza de las variables, función objetivo y restricciones la programación matemática se divide principalmente en programación lineal, entera, mixta y no lineal [36].

7.7 Programación entera

Los problemas de programación entera (ILP) por sus siglas en inglés, son aquellos en los cuales todas las variables de decisión asociadas pertenecen al conjunto de los números enteros positivos (Z^+). Formalmente un problema ILP en su forma canónica puede expresarse como [37]:

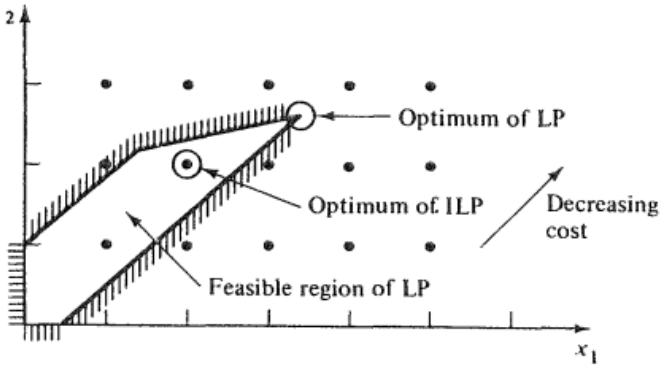
$$\min\{c^t x : x \in X\}$$

$$X = \{x \in Z_+^n : Ax = B\}$$

La necesidad de formular, implementar y solucionar problemas ILP se debe principalmente a que en algunos problemas de programación lineal (LP) las soluciones fraccionadas en las que $\boldsymbol{x}^* : \boldsymbol{x} \in \boldsymbol{R}$, no son factibles. Esto se debe principalmente a que las variables de decisión pertenecen a objetos los cuales sólo tienen sentido si están expresados en cuantías discretas, por ejemplo, recurso humano, número de aviones asociados a una ruta, etc. [30]. De todas maneras existe una gran conexión entre la programación entera, específicamente con la programación lineal (LP), ya que esta última puede ser usada para solucionar instancias del ILP [38], [17] e incluirse en los métodos de solución a través de

relajación de las restricciones del problema si $\boldsymbol{x}$ se considera un número relativamente grande en el que redondear la cifra no afecta significativamente la solución [30]. Aunque en ciertas ocasiones redondear la cifra podría afectar la factibilidad del problema, esto se observa claramente en la figura No. 1 en la cual el óptimo del problema ILP relajado a LP no está dentro de la región factible.

Figura No. 1 Relación entre ILP y LP [39]



Por esta razón se han desarrollado varios métodos y procedimientos para atacar los ILP que garantizan su convergencia y factibilidad.

7.8 Programación binaria

Los problemas de programación lineal binaria (BLP), son un caso particular de los problemas de programación entera, en los cuales todas las variables de decisión deben tomar valores 0 ó 1. La dificultad asociada a los BLP se da cuando existe una gran cantidad de variables de decisión, ya que se amplía considerablemente el espacio solución del problema. Formalmente un BLP se representa de la siguiente manera [40]:

$$\min\{c^t x : x \in X\}$$

$$X = \{x \in \{0,1\} : Ax = B\}$$

La introducción de variables binarias permite incluir decisiones del tipo Sí-No, además de restricciones lógicas a los problemas ILP. Los BLP son comúnmente usados para determinar únicas decisiones que pueden tener variedad de opciones.

Una de las ventajas de la modelación de los BLP son las restricciones lógicas que se pueden asociar al problema, entre ellas [41]:

Exhaustividad:

$$\sum_{i \in K} x_i = 1$$

Esta restricción permite asegurar que sólo se escogerá una variable x que pertenece al conjunto k.
Contingencia:

$$x_i - x_j = 0$$

Esta restricción permite asegurar que si la variable binaria x_i *ó* x_j toma el valor de 1, la variable restante también deberá tomar ese valor (Si $x_i > 0 \rightarrow x_j > 0$).

7.9 Meta-heurística

Los problemas complejos representan usualmente mejor la realidad, en ellos se pueden expresar las relaciones de interacciones y restricciones complejas en las cuales se ven enmarcados los problemas de decisión. Sin embargo, la mayoría de problemas complejos no pueden ser resueltos de manera óptima en un tiempo razonable, por lo que usar métodos aproximados es una buena opción. Comúnmente los problemas reales no requieren de una solución óptima, es suficiente con una solución adecuada a las proporciones del problema [42]. Las meta-heurísticas son procedimientos generales que mezclan estrategias de alto nivel con implementos de búsqueda local [25]. Estos procedimientos tienen inmersas estrategias que permiten escapar de óptimos locales para poder aumentar la probabilidad de encontrar una buena solución. Estos procedimientos genéricos permiten atacar diferentes tipos de problemas de optimización debido a la generalidad de sus lineamientos.

Las meta-heurísticas solucionan instancias de problemas que son considerados como complejos, disminuyendo el espacio solución del problema al explorarlo de manera eficiente. Las meta-heurísticas tienen tres propósitos principales: Resolver el problema más rápido, obtener algoritmos robustos, obtener algoritmos flexibles y fáciles de implementar [42].

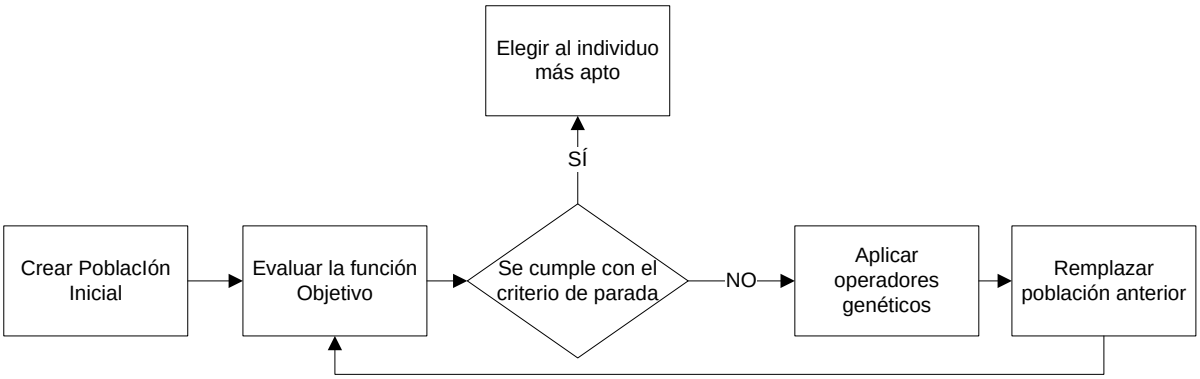
Estos procedimientos han sido el resultado de la interacción de varios campos como: inteligencia artificial, inteligencia computacional, programación matemática, estadística, etc. La mayoría de meta-heurísticas imita los procesos naturales para resolver problemas de optimización complejos. Entre la gran cantidad de meta-heurísticas están: algoritmos genéticos, recocido simulado, colonia de hormigas, colonia de abejas, nube de partículas, búsqueda tabú, entre otros [43].

La importancia de las meta-heurísticas se ha incrementado mucho en los últimos tiempos, cada vez más herramientas y procedimientos son implementados para crear nuevas formas de atacar problemas de naturaleza compleja. En la práctica, estas técnicas participan activamente en la industria manufacturera, industria de servicios, tecnología, finanzas, minería de datos, entre otros; por lo que durante las últimas décadas ha existido un gran aumento en la investigación para el desarrollo de nuevas herramientas y aplicaciones orientadas al desarrollo del campo de las Meta-heurísticas [25]. En el entorno actual, el cual es cada vez más complejo y reactivo por la interacción de la presión económica y la incertidumbre de la demanda, la optimización y el tiempo de solución tienen un papel importante en el desarrollo de ventajas competitivas [42].

7.10 Algoritmos genéticos

Son métodos de búsqueda estocástica que siguen los principios de la genética modelada a través de principios de evolución y de selección natural. El concepto básico de los GA es un sistema que empieza con una población de individuos (soluciones) generadas al azar y que evolucionan mejorando el valor de la función objetivo a través de la implementación de operadores genéticos como cruce o mutación. A cada individuo se le asigna un valor correspondiente a la evaluación de la función objetivo con esa solución. Este método es de gran utilidad cuando el espacio de solución es grande, complejo y cuando se puede determinar fácilmente la función objetivo [43]. De manera muy general se presenta el proceso asociado a los algoritmos genéticos.

Figura No. 2 Proceso general de funcionamiento de los algoritmos genéticos.



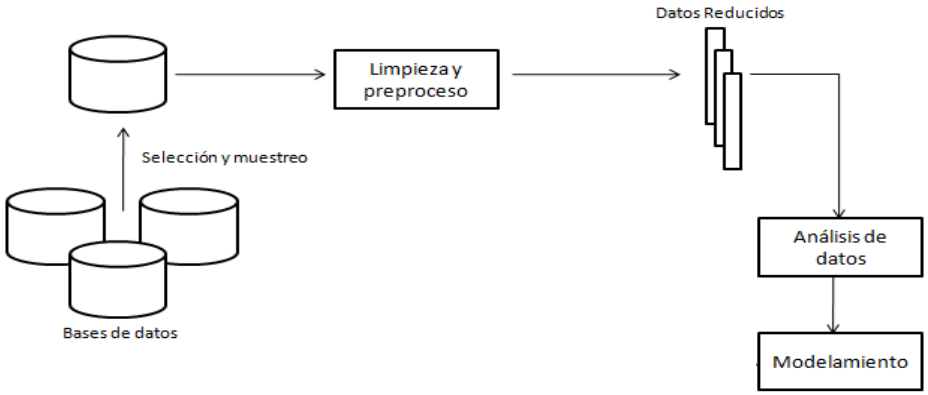
Los criterios de parada están generalmente asociados con el número de generaciones alcanzadas o por el número de generaciones sin mejora en la función objetivo.

8 Metodología

El desarrollo de este proyecto estará alineado con el proceso general de extracción de conocimiento de las bases de datos o *Knowledge data discovery (KDD) process*, el cual se refiere al proceso no trivial de identificar información válida y útil dentro de las bases de datos [44]. El proceso (KDD) juega un rol muy importante en la manera en la que se interactúa con las bases de datos, especialmente en aquellos procesos en los que el análisis y exploración son esenciales. Los procesos KDD han sido ampliamente utilizados durante los últimos años de manera transversal en diferentes áreas del conocimiento, como ingeniería, biología, estadística, modelamiento, optimización, etc., lo que permite que esta metodología sea ampliamente usada para abordar gran variedad de problemas [45].

Los procesos de educación que se llevan en la Universidad generan datos que contienen información relevante del desarrollo académico de los estudiantes, que, por lo general sólo es enfocada hacia las auditorías de calidad [3]. Estos datos generados por los programas académicos requieren protagonizar un papel más influyente en los procesos educativos de la institución. En particular, para este proyecto se ha definido la siguiente metodología, basándose en los pasos propuestos por *Usamma Fayyad* [46], a través de los cuales se pretende enmarcar de manera general el desarrollo de este proyecto de investigación.

Figura No. 3, Metodología general de desarrollo del proyecto



8.1 Selección de datos

Consiste en recopilar sólo la información que sea relevante para el desarrollo del proyecto. Se catalogarán las bases de datos como válidas sólo si están dentro del rango de tiempo en el cual el sistema de créditos fue implementado.

8.2 Limpieza y pre-proceso

Se refiere al proceso de depuración de la base de datos, en la cual se eliminan los registros que tengan información incoherente y que no aporten valor al modelo. Por la naturaleza de la base de datos la mayor parte de este proceso consiste en apartar los registros, homogeneizar los nombres de las asignaturas y completar información.

8.3 Análisis de datos

Comprende dos fases internas

8.3.1 Caracterización de variables

Consiste en identificar el comportamiento propio de cada una de las variables involucradas en el proceso, para poder así determinar la manera más adecuada de incluirlas dentro del modelo.

8.3.1.1 Asignaturas

A cada una de las asignaturas y grupos de asignaturas se les debe aplicar las siguientes herramientas de la estadística descriptiva:

- Promedio
- Mediana
- Desviación estándar
- Cuartil no. 1 y no. 3
- Percentil no. 5 y no. 90
- Coeficiente de Kurtosis
- Coeficiente de Asimetría
- Porcentaje de notas sobre 3
- Promedio por periodo académico
- Histograma
- Caracterización de la distribución de probabilidad asociada a cada asignatura
- Prueba de bondad de ajuste para cada distribución

8.3.1.2 Números de clase

Para analizar el comportamiento de los números de clase de cada una de las asignaturas de NFF y ENF es necesario utilizar las siguientes herramientas de la estadística descriptiva.

- Gráficos de control de los números de clases de una asignatura.
- Ranking de las asignaturas con mayor número de clases fuera de los límites de control.
- Gráfico de cajas de cada número de clase de una asignatura para un periodo en particular.

8.3.1.3 Asignaturas re-cursadas

Con el fin de conocer más el comportamiento de los resultados de cursar una asignatura más de una vez se debe realizar el siguiente análisis para cada asignatura que tuviera estudiantes que la hayan re-cursado al menos una vez.

- Agrupar el número de veces que se re-cursa una asignatura e indicar el promedio para cada una de ellas
- Detallar los cambios en las distribuciones de probabilidad asociadas al número de veces que se repite una asignatura.
- Graficar el número de estudiantes que re-cursan una asignatura en particular.
- Identificar el posible número de estudiantes que se retiran del programa académico según cada asignatura.

8.3.1.4 Requisitos

Para facilitar el acompañamiento de los consejeros académicos en el proceso de selección de asignaturas por parte del estudiante es de gran utilidad construir un proceso automatizado que permita seleccionar las asignaturas que se deben aprobar para estar en condiciones académicas de cursar una asignatura en particular y también seleccionar el conjunto de asignaturas candidatas a inscribir según el histórico académico de cada estudiante en particular.

8.3.1.5 Interacción entre variables

La precedencia de cursos en el programa educativo puede servir como escenario para el análisis de correlaciones entre cada asignatura y los resultados del conjunto de asignaturas cursadas hasta el periodo anterior. Definiendo $\boldsymbol{G}$ al conjunto de asignaturas vistas anteriormente, y a $\boldsymbol{x}$ la asignatura a analizar, se pretende identificar el subconjunto $\boldsymbol{G'} \in \boldsymbol{G}$, que tiene incidencia en los resultados de $\boldsymbol{x}$, ya sea de manera directa o a través de su interacción con otras variables del conjunto. Para esto, es necesario implementar los siguientes procesos.

- Transformar toda la base de datos y orientarla hacia el registro histórico de cada estudiante.
- Obtener la matriz de correlación con el coeficiente de correlación lineal Pearson entre asignaturas.
- Realizar una prueba de hipótesis de significancia del coeficiente de correlación lineal Pearson
- Identificar las 10 asignaturas que tienen más influencia en cada una de las asignaturas de NFF y ENF.
- Crear un gráfico que represente de manera clara la relación entre cada pareja de asignaturas.

8.4 Modelamiento

Para poder representar el problema en un contexto matemático que permita explotar todos los avances desarrollados en problemas combinatorios, es necesario identificar apropiadamente las siguientes características asociadas al modelo.

8.4.1.1 Función objetivo

Se debe representar una estimación del valor esperado del promedio ponderado acumulado $\boldsymbol{P}$ según el resultado de las calificaciones históricas y la estimación de las calificaciones sugeridas. Si se representa al conjunto total de asignaturas como $\boldsymbol{A}$, y $\boldsymbol{G}$ el conjunto de asignaturas cursadas hasta el momento, se pretende estimar $\boldsymbol{E}[\boldsymbol{P}|\boldsymbol{G}]$, $\boldsymbol{G} \in \boldsymbol{A}$. Este estadístico permitirá evaluar cada una de los posibles subconjuntos $\boldsymbol{s}$, $\boldsymbol{s} \in \{\boldsymbol{A} - \boldsymbol{G}\}$ que cumplan con las restricciones pertinentes. Además se representará el aporte de las asignaturas al plan de estudios, en particular, si se considera a la configuración de las asignaturas para cada periodo académico como un grafo dirigido $\mathbf{G}(\mathbf{N}, \mathbf{A})$ en el que $\boldsymbol{N}$ representa a los nodos o asignaturas, y $\boldsymbol{A}$, son todos los arcos $\boldsymbol{i}, \boldsymbol{j}$ que expresan la relación de dependencia dirigida de $\boldsymbol{i}$ a $\boldsymbol{j}$; la contribución de una asignatura específica al cumplimiento del plan de estudios será proporcional al número de nodos $\boldsymbol{A}$ son precedidos por ella.

8.4.1.2 Restricciones

Se debe definir las limitaciones que reduzcan el espacio factible del total subconjuntos de asignaturas posibles a inscribir. Para ello entonces se requiere analizar las restricciones asociadas a:

- Cantidad máxima de créditos según la modalidad de matrícula.
- Cantidad mínima de créditos que el estudiante está dispuesto a cursar.
- Cantidad mínima y máxima de créditos de asignaturas electivas y complementarias.
- Precedencia de asignaturas.

8.5 Métodos de solución

Dada la complejidad del problema es necesario identificar un método de solución eficiente que permita resolver el modelo en un tiempo de ejecución aceptable. Con la identificación de los tipos de soluciones obtenidas para las diferentes instancias del KN que pueden adaptarse a las particularidades del problema de selección de asignaturas, se propone el siguiente método de solución.

- Relajación lineal de la función objetivo.
- Solución del problema relajado a través de métodos exactos.
- Incluir el anterior resultado como solución inicial en un algoritmo genético. Para la construcción del algoritmo genético es fundamental describir claramente el cromosoma asociado, la función objetivo y los operadores genéticos.

8.6 Procedimiento detallado de la metodología

En la figura No. 2 se encuentra detallado de manera gráfica la metodología de análisis de datos y en la figura No. 3 la metodología para la selección de asignaturas.

Figura No. 4 Procedimiento detallado de la metodología de análisis de datos.

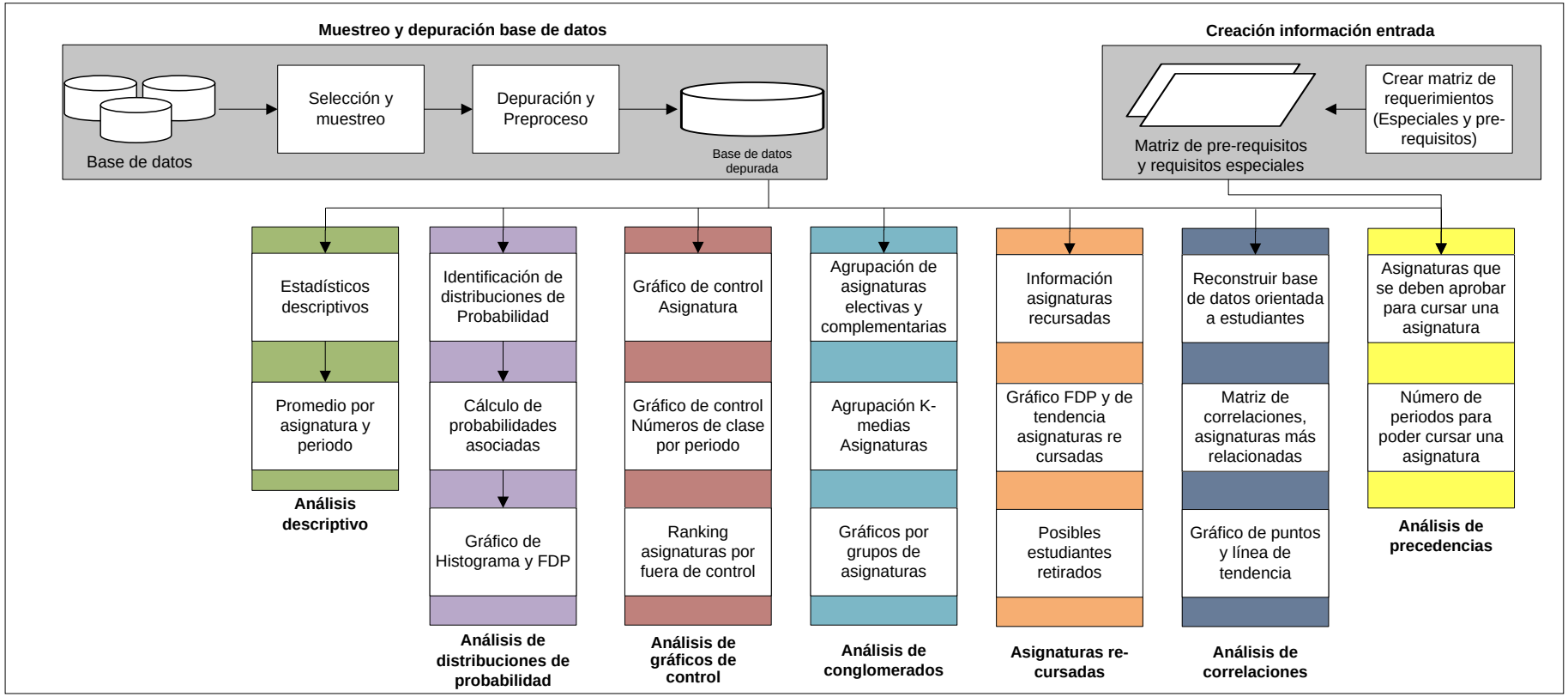
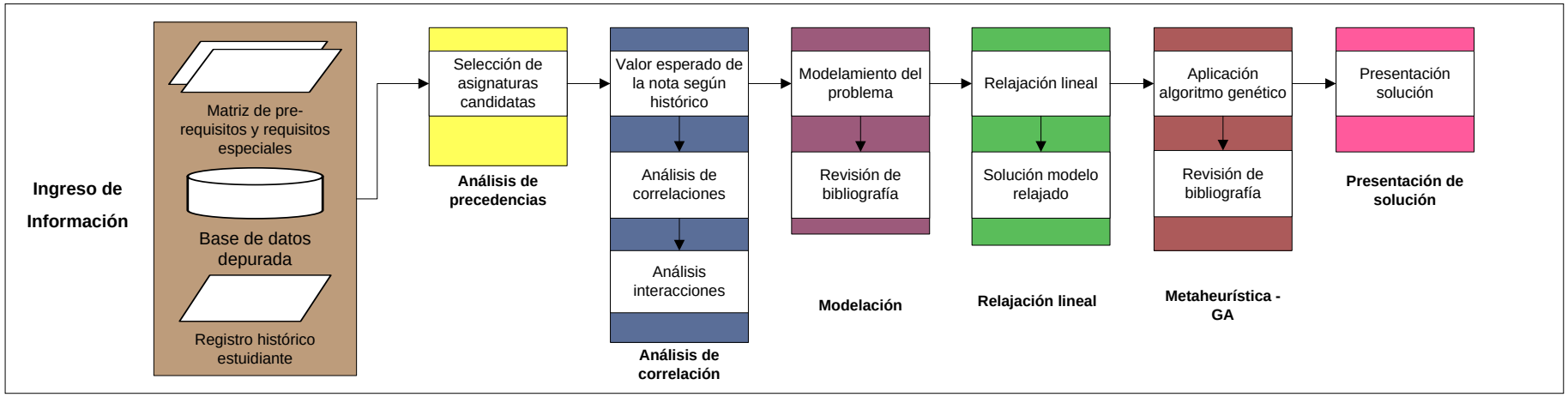


Figura No. 5 Metodología detallada para el procedimiento de selección de asignaturas.



9 Implementación de la metodología

A continuación se presentan los diferentes apartados correspondientes a la metodología propuesta.

9.1 Requerimientos de información

La metodología propuesta está dividida principalmente en dos secciones: Análisis de la base de datos de registros académicos y el procedimiento para la selección de asignaturas a cursar por un estudiante. A continuación, se describe los requerimientos de información necesarios para llevar a cabo cada una de las instancias de esta metodología.

9.1.1 Análisis de base de datos

El proceso para guiar el análisis y caracterización de las asignaturas requiere de las siguientes entradas de información.

9.1.1.1 Base de datos académicos

La base de datos inicial es la resultante de solicitar la siguiente consulta en la herramienta SIU

SAE Producción → Catálogo de consultas SAE → Módulo Registro Estudiantil →División Situaciones Académicas y Notas → Histórico Notas por Programa

Es preciso contar con bases de datos robustas que representen fielmente la situación que se desee analizar. Entre algunos de los criterios que se deben tener en cuenta al momento de seleccionar los periodos académicos que conformarán la base de datos están:

- Periodos académicos en los que ya esté funcionando con normalidad el sistema de créditos académicos
- Periodos académicos en los que el plan de estudios del programa académico pueda considerarse como una aproximación al plan de estudios que esté vigente en la actualidad.
- Periodos académicos en los cuales la universidad no haya tenido una restructuración importante que afecte la fiabilidad de los registros.
- Periodos académicos en los cuales se disponga de la información necesaria para usar esta metodología.

Para que sea posible realizar el análisis de manera adecuada es necesario que la base de datos esté orientada por registros y contenga los siguientes campos nombrados estrictamente (distinguiendo mayúsculas) como se presentan a continuación, de igual manera se debe respetar el orden estipulado.

- **ESTUDIANTE:** Número de identificación del estudiante, en este caso el ID.
- **PERIODO:** Periodo académico numérico (Un mayor número debe indicar un periodo más reciente).
- **IDCURSO:** Número de identificación de la asignatura.
- **MATERIA:** Nombre de la asignatura.
- **Nclase:** Número de clase del curso perteneciente a la asignatura.
- **CREDITOS:** Número de créditos académicos de la asignatura.
- **NOTA:** Calificación obtenida por el estudiante en la asignatura en ese periodo (el separador decimal debe ser coma “,”).

- **ORGACA:** Organización académica responsable de impartir la asignatura.
- **NATURALEZA:** Naturaleza de la asignatura, Núcleo de Formación Fundamental (NFF), Énfasis (ENF), Opción complementaria (CP), Electivas (ELE).

9.1.1.2 Matriz de Pre-requisitos

Debe ser una matriz $\mathbf{NxN}$, donde $\mathbf{N}$ es el número de asignaturas que conforman el núcleo de formación fundamental y énfasis. Esta matriz $\mathbf{M}$ debe cumplir que:

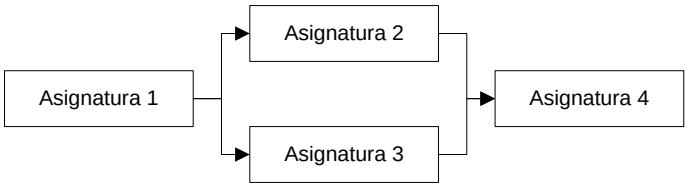
$$\mathrm{M_{ij}} = \begin{cases} 1, & Asignatura\ i\ es\ requisito\ para\ cursar\ j \\ 0, & en\ otro\ caso \end{cases}$$

Si la posición $\boldsymbol{M_{12} = 1}$ entonces la asignatura 1 debe ser aprobada para estar en condiciones académicas de cursar la asignatura 2. Por definición de esta matriz, cada uno de los componentes $\boldsymbol{M_{ii}}$ deben ser 0. Si llegase a existir un correquisito entre una pareja de asignaturas, entonces ambas deberán compartir sus pre-requisitos. A continuación, se puede detallar una matriz de requisitos de 4 asignaturas en donde la asignatura 1 es pre-requisitos de la 2 y 3, mientras que las asignaturas 2 y 3 son pre-requisitos de la 4. En el anexo digital llamado "Matriz Requisitos" se encuentra disponible la Matriz de Pre-requisitos del programa de ingeniería Civil de la Pontifica Universidad Javeriana – Sede Bogotá, creado a partir de la información dispuesta en la página web del programa académico en el primer día del periodo 1210.

Tabla No. 2. Ejemplo reducido de matriz de requisitos.

	Asignatura 1	Asignatura 2	Asignatura 3	Asignatura 4
Asignatura 1	0	1	1	0
Asignatura 2	0	0	0	1
Asignatura 3	0	0	0	1
Asignatura 4	0	0	0	0

Figura No. 6. Visión de grafo del ejemplo reducido de matriz de requisitos.



La figura No. 6 representa una manera diferente de expresar los pre-requisitos entre asignaturas. Para este caso cada una de las asignaturas representa un nodo y la dirección de las flechas indican la relación de dependencia.

9.1.1.3 Requisitos especiales

Algunas asignaturas requieren el cumplimiento de algunos requisitos particulares para que se puedan cursar. La metodología desarrollada contempla los requisitos de número de créditos aprobados y cumplimiento de lengua extranjera. Esta información debe presentarse en una tabla con el siguiente formato.

Tabla No. 3. Estructura matriz de requerimientos especiales.

Asignaturas	Requisitos de créditos aprobados	Requisito de lengua extranjera
Nombre asignatura	Número de créditos	sí/no

La matriz de requerimientos especiales del programa de Ingeniería Civil de la Pontificia Universidad Javeriana se encuentra disponible en los anexos digitales bajo el nombre "Requisitos Adicionales de Asignaturas", la cual fue construida a partir de la información dispuesta en la página web del programa académico en el primer día del periodo 1210.

9.1.2 Procedimiento de selección de asignaturas

Consiste en la automatización del proceso de sugerencia de lo conjuntos de asignaturas que se deben tener en cuenta para el siguiente periodo.

9.1.2.1 Registro histórico del estudiante

La metodología que se desarrolla en este apartado requiere la información académica histórica de un estudiante. En el estado actual del desarrollo de esta metodología aún no es posible integrar la información relacionada con las asignaturas re-cursadas en la estimación de las calificaciones esperadas, por lo que en este registro, sólo debe aparecer la última vez que el estudiante cursó la asignatura, ya sea que la haya aprobado o reprobado. El nombre de cada asignatura debe ser exactamente igual a como están dispuestas en la base de datos.

Además de contener el histórico de las asignaturas cursadas, esta tabla indica la opción complementaria elegida por el estudiante y el estado del requisito de idioma extranjero. Esta información debe estar construida de la siguiente manera:

Tabla No. 4. Estructura matriz registro histórico de un estudiante.

Cursadas	**Nota**	**Complementaria**
Requisito idioma extranjero	Sí/No	Organización académica responsable
Nombres de las asignaturas	Calificación de cada asignatura	

Si el estudiante no tiene aún el requisito de idioma extranjero, en la casilla nota debe estar la palabra 'No', mientras que para indicar que el estudiante no tiene una complementaria definida se debe escribir en ese campo 'NA'. En los archivos anexos digitales se encuentra un ejemplo del registro histórico de un estudiante bajo el nombre de "Entrada Datos Estudiante".

En cada uno de los ingresos de información es necesario que haya exactitud y coherencia entre todas las palabras implementadas, además de respetarse los nombres de cada una de los componentes de las tablas incluyendo la manera en la cual están escritos (espacios, tildes, mayúsculas).

9.2 Entorno para el desarrollo del proyecto

A continuación se menciona las herramientas de software que se proponen para el uso de las instancias de la metodología propuesta.

9.2.1 Ingreso de información

Para la depuración de la base de datos y creación de las tablas de información de entrada indicadas en la sección 9.1.1 se recomienda usar Microsoft Excel.

9.2.1.1 Microsoft Excel

Es una aplicación de hoja de cálculo que viene con el paquete ofimático de *Microsoft Office*. Excel por lo general es una herramienta orientada a manejo de datos contables, financieros y estadísticos pero que puede ser usada en una gran variedad de campos [47]. Excel soporta una de las herramientas más funcionales para la automatización de proceso, las macros, las cuales

son instrucciones secuenciales que se pueden realizar de manera automática y ejecutar a través de un solo comando. De igual manera también permite la validación de los campos de entrada para asegurar una unificación de los conceptos en las bases de datos.

9.2.1.2 Ventajas identificadas

Las ventajas que tiene Excel en el manejo de la información relacionada con las bases de datos y registros de los estudiantes para la aplicación de esta metodología son principalmente:

Universalidad: Excel viene preinstalado en todos los ordenadores que tienen *Windows*, y de igual manera es posible instalarlo en los diferentes sistemas operativos. Por lo que se asegura que la mayoría de los usuarios de esta metodología tendrán al alcance esta herramienta ofimática.

Facilidad de manejo: Excel es una herramienta que utiliza comandos muy sencillos y que tiene predefinidas varias funciones útiles. Además, debido a su universalidad se podría suponer que los usuarios y beneficiaros de esta metodología tienen un manejo básico de esta herramienta.

Entorno gráfico: Excel tiene una GUI muy amigable, por lo que no sería problemático que los usuarios y beneficiarios de esta metodología que no hayan tenido un acercamiento a esta herramienta se les instruya de manera rápida en los temas pertienentes para la recopilación de información necesaria para aplicar la metodología (Conceptos de celdas, ingreso de información, grabar archivos).

Para conocer más sobre el manejo de Microsoft Excel se recomienda seguir la guía propuesta por *Velosa*, en el siguiente enlace verificado el día 10-Abr-2013:

http://tinyurl.com/c49dooy

9.2.2 Procesamiento de datos

En el procesamiento de datos necesario para el análisis de asignaturas y el desarrollo de la modelación del proceso recomendación se sugiere utilizar *R statistics*.

9.2.2.1 R statistics

Es un entorno de programación que contiene integrado programas e instancias para la manipulación de datos, simulación, cálculos y gráficos. R tiene doble naturaleza, de programa y lenguaje de programación basado en un dialecto del lenguaje S creado por AT&T Bell [48], el cual se distribuye de manera gratuita bajo los términos de la GNU.

R es uno de los entornos que más se está desarrollando hoy día. Posee una gran cantidad de funciones que se incluyen al momento de instalar el programa, y ofrece un gran número de paquetes especializados muy importantes y vigentes que pueden conseguirse en la web, [49]. Además, es posible encontrar en la página de paquetes de los colaboradores, funciones especializadas en diversos campos de la estadística y matemática. Uno de los atractivos de R es que incluye un lenguaje de programación bien desarrollado, simple y efectivo, que admite condicionales, bucles, funciones recursivas.

9.2.2.2 Ventajas inidentificadas

Para el desarrollo de esta metodología las principales ventajas que ofrece R sobre los demás paquetes estadísticos son principalmente:

Software libre: R se puede descargar de manera gratuita y legal desde cualquier ordenador con conexión a internet, lo cual es una grande ventaja comparada con las licencias pagas de los demás software de estadística como *SPSS, Minitab, SAS*, etc.

Colaboración global: Debido a la naturaleza libre de R, existen amplios y avanzados desarrollos en la creación de funciones específicas para R en diversos campos de la estadística y matemática, tales como: Análisis de Conglomerados, Gráficos avanzados, Programación mixta, etc.

Flexibilidad: La doble naturaleza de R permite combinar varias funciones de programación y estadística y aplicarlas a programas específicos que se adapten de manera eficiente a la estructura de las bases de datos. R brinda entonces el potencial para realizar casi cualquier análisis estadístico y matemático a cualquier estructura de base de datos en la manera que el usuario lo desee, mientras que los demás paquetes estadísticos tienen normas rígidas para el uso de las funciones preinstaladas, su método de aplicación y presentación de resultados

Por otra parte la gran debilidad de R es su barrera de entrada, ya que no cuenta con una interfaz de usuario agradable y podría parecer demasiado complicado para los no especialistas.

Para un conocimiento más profundo que incluye el proceso de instalación de esta herramienta de análisis de datos, se recomienda leer el manual provisto en el siguiente enlace, verificado el día 10-Abr-2013:

http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf

9.2.2.3 Paquetes utilizados

Gracias al desarrollo conjunto y constante actualización que existe de los diferentes paquetes estadísticos y matemáticos, este proyecto está soportado en algunas rutinas desarrolladas y testeadas, las cuales se mencionan a continuación:

Rcmdr: Crea una interfaz gráfica que facilita le ejecución de comandos básicos y la visualización de la información de salida y errores de sintaxis de programación [50].

Rgenoud: Es una función creada en R de algoritmos evolucionarios de búsqueda basado en derivadas creado por *Walter R, Mebane y Jasjeet Singh y Sekhon*, aunque de igual manera, también es posible usarlo en problemas de optimización donde las derivadas no existen [51]. Esta función permite utilizar los métodos desarrollados de algoritmos genéticos definiendo únicamente los parámetros cromosoma y función objetivo, no obstante, también permite la manipulación de los demás parámetros asociados a los AG.

Rglpk: Es una función que brinda una interfaz para utilizar GLPK en R [52].

9.3 Estructura de presentación de las funciones creadas en la metodología

Para facilitar la visualización las funciones más relevantes creadas en R para esta metodología, cada una de ellas será presentada bajo la siguiente estructura.

- **Función:** Nombre de la función.
- **Hipervínculo:** Enlace que dirige hacia el código desarrollado de la función
- **Sintaxis:** Indica cómo debe ser invocada la función.
- **Parámetros:** Qué requisitos de información tiene la función.
- **Función requerida:** Función que se debe ejecutar antes.
- **Objetivo:** Menciona cuál es el propósito de la función.
- **Salida:** Qué arroja la función.
- **Precondiciones:** Supuestos bajo los cuales está creada la función.
- **Diagrama de flujo:** Muestra de manera gráfica las secuencias de la función.
- **Ejemplo:** Presenta un ejemplo del uso de la función
- **Visualización salida:** Muestra una impresión de pantalla de los productos de salida.

9.3.1 Ejemplo para el desarrollo de la presentación de la metodología

Para facilitar el entendimiento de la metodología se trabajará en cada una de las funciones un ejemplo relacionado con el siguiente registro histórico académico de un estudiante y la asignatura **Ecuaciones diferenciales,** la cual tiene 3 créditos y está identificada con el código 001300**.**

Tabla No. 5. Ejemplo de registro histórico de un estudiante para detallar el desarrollo de la metodología.

Aprobadas	Nota	Complementaria
Requisito Idioma Extranjero	NO	DPT-LENG
Cálculo Diferencial	3.2	
Introducción Ing Civil	4.4	
Expresión Gráfica y Geometría	4.3	
Química de Materiales	4.1	
Inglés 1	4.7	
Epistemología de la Ingeniería	4.2	
Álgebra Lineal	3.3	
Baloncesto, Iniciación.	5.0	

Con este ejemplo se puede detallar (según la información de la base de datos) que el estudiante aún no tiene el requisito de idioma extranjero cumplido y tiene 20 créditos aprobados, con un promedio ponderado de 4.3. También es posible observar que tiene una asignatura complementaria de 3 créditos y una asignatura electiva de 2 créditos.

9.4 Caracterización, pre-proceso y depuración de la base de datos

9.4.1 Características generales de la base de datos

El presente trabajo de grado está desarrollado en la base de datos provista por la carrera de Ingeniería Civil, la cual incluye aproximadamente 21.000 registros que van desde el primer semestre del año 2007 hasta el primer semestre del año 2012.

9.4.2 Consenso de términos

Para que el análisis de la base de datos genere información relevante y fiable es necesario realizar una estandarización de los nombres de las asignaturas y sus organizaciones académicas. Para ello se recomienda seguir estos cuatro pasos

9.4.2.1 Identificación de registros con errores ortográficos

Añadir una columna a la base de datos y con la ayuda de la función CONTAR.SI contar todos los registros que tengan el nombre de la asignatura estipulado en esa fila. Para una base de datos de 1000 registros que empieza desde la segunda fila y que el campo asignatura está en la cuarta columna, el nuevo campo añadido debe tener una formulación que debe ser similar a CONTAR.SI($D$2:$D$1000;$D2).

Después se procede a copiar sólo los valores de la columna creada junto a la columna asignaturas, por último con ayuda de la función eliminar duplicados ubicada en la pestaña datos, eliminar todas las parejas de estas dos columnas que tengan datos repetidos. De esta manera una misma asignatura deberá tener el mismo valor en la nueva columna creada, por lo que un número diferente significará que la escritura de esta asignatura no es homogénea.

9.4.2.2 Identificación de registros con diferencias de letras mayúsculas y minúsculas

Después de realizar la corrección de las asignaturas con disparidades ortográficas se deberá crear una nueva columna en donde se hará el mismo procedimiento del paso anterior pero diferenciando también las letras mayúsculas y minúsculas. Esta nueva columna debe tener una formulación similar a {=+SUMA(SI(IGUAL($D$2:$D$1000;$D2);1;0))}. En este caso nos hemos soportado en la formulación matricial para poder invocar la función IGUAL, se debe tener en cuenta que para que las fórmulas matriciales funcionen es necesario oprimir CTRL+SHIFT+ENTER al momento de ingresar la ecuación.

9.4.3 Campo organización académica (ORGACA)

Es probable que en la base de datos inicial existan varios registros que no contienen el campo organización académica, el cual representa el departamento que imparte la asignatura. Este campo es imprescindible para la realización de la metodología por lo que a continuación se presentan algunas sugerencias para completar este campo.

9.4.3.1 Identificación de las asignaturas sin organización académica

Crear una copia de las columnas IDCURSO y ORGACA de la base de datos, en una hoja aparte con la herramienta Quitar Duplicados disponible en la pestaña Datos, eliminar los registros repetidos de ambos campos. Acto seguido se debe hacer lo posible por llenar todos los campos vacíos investigando con las herramientas disponibles (SIU, Directores de carrera) cuál organización académica tiene la responsabilidad sobre esa asignatura.

9.4.3.2 Diligenciar la organización académica

Ahora es preciso copiar las organizaciones académicas de los datos duplicados y diligenciados del paso anterior en los campos correspondientes a las asignaturas que por una u otra razón no las contienen. Para este procedimiento utilizamos la función BUSCARV. Suponiendo que se tiene un total de 100 asignaturas y 1000 registros, y el código asignatura está en la tercer columna, la formulación que debe ir bajo el título ORGACA y la cual se debe replicar para todos los registros tiene que ser similar a BUSCARV($C2; $AA$1:$AB:100; 2; FALSO). De esta manera se busca la organización académica correspondiente a la asignatura en C2 en la matriz AA1:AB100, la cual contiene la copia de IDCURSO y ORGACA y procede a llenar el campo vacío con la información obtenida en el paso anterior.

9.5 Funciones creadas en el entorno de programación R statistics

A continuación se presentan las funciones creadas en el entorno de programación R para el desarrollo de cada uno de los apartados de la metodología propuesta.

9.5.1 Apartado estadística descriptiva

Este apartado está enfocado al desarrollo de funciones que aporten información acerca de cada una de las asignaturas dispuestas en la base de datos con el fin de poder caracterizarlas y obtener una idea general del comportamiento de sus calificaciones.

9.5.1.1 Función: Información estadística básica de las asignaturas

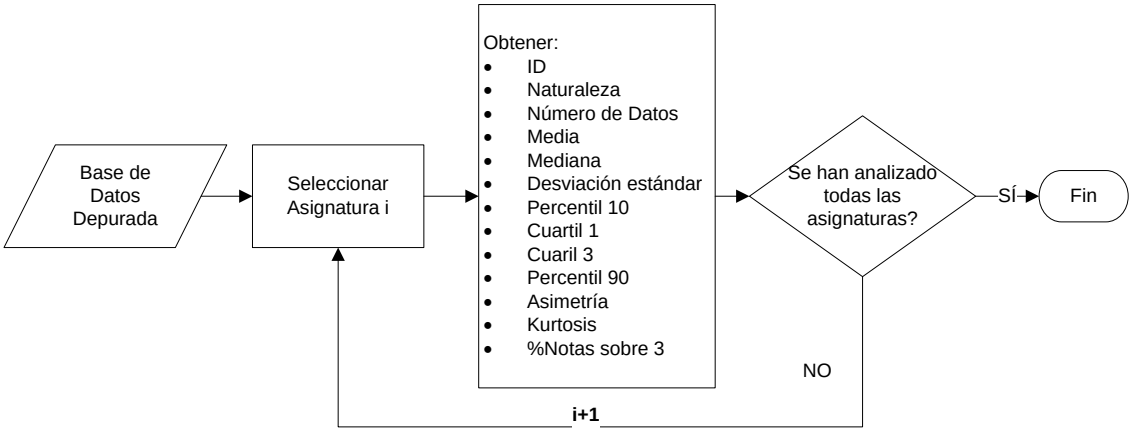
Apartado: Estadística descriptiva
Nombre de la función: Información estadística básica de las asignaturas
Hipervínculo: Código – Invocación
Sintaxis: MatrizEsta(BaseDatos)
Parámetros: Base de datos académica depurada.
Salida: Tabla con información general y estadísticos de tendencia central de todas las

asignaturas.

Objetivo: Servir como método de consulta para caracterizar el comportamiento de las calificaciones de una o varias asignaturas.

Precondiciones: El número de datos válidos de cada asignatura es suficiente para caracterizarla según sus estadísticos de tendencia central.

Diagrama de Flujo:



Ejemplo:

//MatrizEsta(BaseDatos)

//showData(na.omit(MatrizEstadisticosDescriptivos))

Visualización Salida:

	ID	Naturaleza	N	Media	Mediana	SD	Percentil 10	Percentil 25
Desastres en Ingeniería	22721	ELE	335	3.93	4.2	1.04	3	3.7
Descub. el Potencial Interno	18874	ELE	9	4.8	5	0.32	4.42	4.7
Descubriendo Japón	23511	CP	4	3.95	3.95	0.21	3.76	3.85
Desordenes alimentarios	3233	ELE	2	4.2	4.2	0.42	3.96	4.05
Dibujo Ingeniería de Producto	8151	ELE	5	4.08	4.3	0.65	3.4	4
Dinámica Estructural	4007	ENF	43	4.06	4.1	0.43	3.7	3.8
Discapacidad y Sociedad	21901	ELE	4	3.75	3.7	0.17	3.63	3.68
Diseño de Fundaciones	4008	NFF	225	3.67	3.8	0.74	3.1	3.3
Diseño en Concreto	3183	NFF	213	3.99	4	0.67	3.4	3.8
Diseño y Arte en Cerámica	105	ELE	6	4.7	4.6	0.2	4.55	4.6
Drenaje-Tratamiento Aguas R	4009	ENF	15	3.99	4	0.57	3.5	3.55
Ecología de Campo	1816	ELE	6	3.85	3.75	0.21	3.7	3.7
Ecología para Comunicadores	23552	ELE	4	4.1	4.05	0.14	4	4
Ecología para todos	21845	ELE	19	4.21	4.3	0.5	3.58	4.1
Ecología Urbana	20025	ELE	2	4.7	4.7	0.42	4.46	4.55
Ecología y la Empresa	23557	ELE	5	4	4	0.19	3.84	3.9
Ecoteología del Desarrollo	2456	ELE	3	2.63	3.5	1.86	1.1	2
Ecoteología y Sentidos de vida	2457	ELE	2	4.35	4.35	0.21	4.23	4.28
Ecoturismo	1800	ELE	2	4.4	4.4	0	4.4	4.4
Ecuaciones Diferenciales	1300	NFF	417	3.16	3.1	0.73	2.2	3

9.5.1.2 Función: Promedio por periodo y asignatura

Apartado: Estadística descriptiva

Función: Promedio por periodo y asignatura

Hipervínculo: Código - Invocación

Sintaxis: MatrizAveMa(BaseDatos)

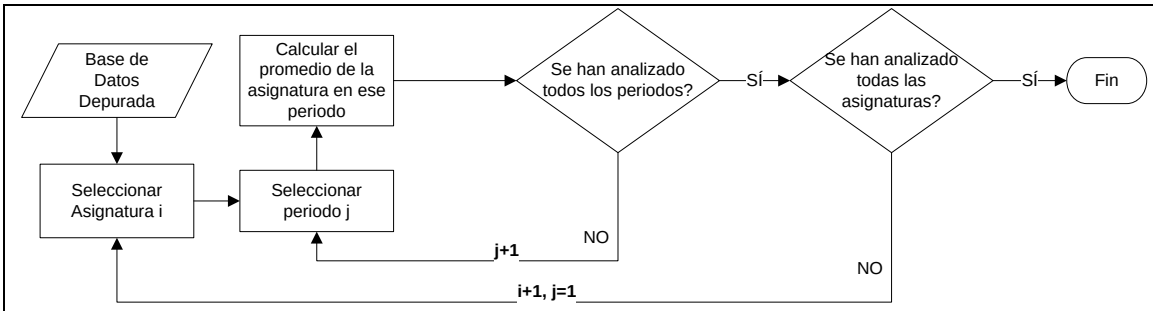
Parámetros: Base de datos académica depurada.

Salida: Tabla con el promedio de calificaciones de cada asignatura en cada periodo académico.

Objetivo: Permitir la evaluación general de la tendencia de una asignatura analizando el comportamiento del promedio de calificaciones en el tiempo.

Precondiciones: El número de datos de cada asignatura es suficiente para considerar a la media aritmética como una medida válida de tendencia central.

Diagrama de Flujo:



Ejemplo:

//MatrizAveMa(BaseDatos)

//showData(na.omit(MatrizPromedioMaterias))

Visualización Salida:

	710	730	810	830	910	930	1010	1030	1110	1130	1210
Álgebra Lineal	3.33	2.96	3.41	3.40	3.33	3.60	3.32	3.54	3.39	3.24	3.27
Análisis Estructural	3.12	3.30	3.38	3.57	3.90	4.09	3.51	4.01	4.26	3.70	3.42
Análisis Numérico	4.50	3.49	4.17	3.26	4.67	4.01	3.38	3.23	3.50	3.74	3.74
Cálculo Diferencial	3.29	3.08	3.10	3.23	3.14	2.93	2.81	3.02	2.91	3.22	2.90
Cálculo Integral	3.09	3.16	3.22	3.41	3.17	3.10	3.02	3.13	3.09	3.08	3.22
Cálculo Vectorial	2.91	2.80	3.15	2.99	2.43	3.20	3.18	3.27	2.99	3.29	2.63
Cardio, acondicionamiento.	4.60	3.40	4.60	4.47	3.55	4.80	4.90	5.00	3.43	3.65	4.35
Constitución y Derecho Público	5.00	4.61	4.33	4.34	4.57	4.67	4.38	4.46	4.65	4.58	4.57
Diseño de Fundaciones	3.67	3.19	2.91	3.75	3.64	3.81	3.85	3.83	4.11	3.65	3.98
Diseño en Concreto	4.47	3.88	3.74	4.15	3.90	4.43	3.63	3.69	4.65	4.16	3.89
Ecuaciones Diferenciales	2.84	3.17	3.29	3.08	3.05	3.15	2.97	3.18	3.10	3.32	3.42

9.5.2 Apartado distribuciones de probabilidad

Este apartado está dirigido a la estimación de las distribuciones de probabilidad asociadas a las calificaciones de las asignaturas.

9.5.2.1 Función: Matriz de probabilidad por asignatura

Apartado: Distribuciones de probabilidad
Función: Matriz de probabilidad por asignatura.
Hipervínculo: Código – Invocación
Sintaxis: MatrizProMa(BaseDatos)
Parámetros: Base de datos académica depurada.
Salida: Tabla con el tipo de distribución –si aplica- (Beta, Discreta), sus parámetros y la significancia del estadístico de Kolmogorov – Smirnov para cada asignatura y distribución estimada.
Objetivo: Obtener una idea más detallada del comportamiento de las calificaciones de una asignatura y servir como parámetro de entrada para algunas funciones posteriores.

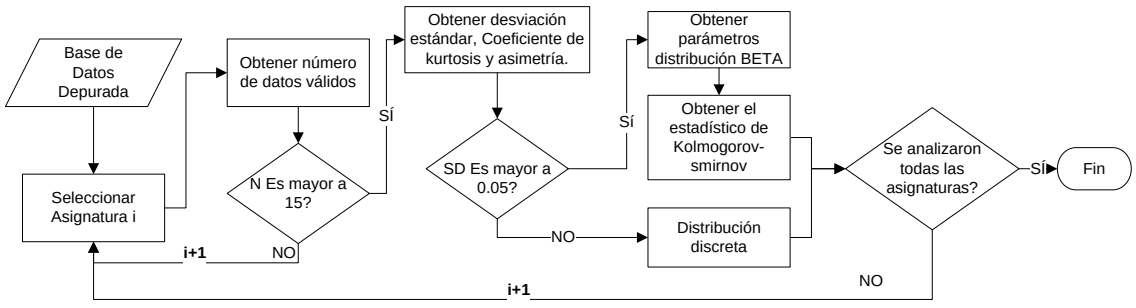
Precondiciones: El comportamiento de las calificaciones de las asignaturas tiene una asimetría negativa, positiva, o cercana a cero dependiendo de los resultados del desempeño de los estudiantes. Estas tendencias han sido modeladas a través de la Distribución Beta, la cual permite simular este tipo comportamientos, además de ofrecer curvas suaves y flexibles, [53], [54]. De igual manera también se asume que aquellas asignaturas que tengan una desviación estándar significativa, en nuestro caso arbitrariamente mayor a 0.05 siguen una distribución Beta. Aquellas asignaturas con menor desviación no pueden ajustarse a este tipo de distribución debido a que tienen concentradas sus notas alrededor de un solo valor. Para este caso se usará únicamente la distribución de frecuencias relativas de las calificaciones de esa asignatura como su distribución de probabilidad. De igual manera sólo se obtendrá una distribución de probabilidad a aquellas asignaturas que tengan al menos 15 datos, ya que se consideran necesarios para garantizar el ajuste de la distribución en gran parte de su rango [55].

Existen varios estadísticos para estimar qué tan bien se ajusta un conjunto de datos a una distribución específica, en el marco teórico del presente trabajo de grado se mencionaron los tres métodos más comunes y ampliamente utilizados. De los métodos presentados es preciso usar el estadístico de Kolmogorov – Smirnov debido a que es más apropiado para distribuciones continuas además de no ser una prueba específica para una sola distribución; como por ejemplo, la prueba de Shapiro Wilk se recomienda para pruebas de normalidad, [32].

Según el análisis realizado, la mayoría de asignaturas en las que la significancia del estadístico de Kolmogorov – Smirnov indica que la distribución de probabilidad no se ajusta a los datos de las calificaciones de una asignatura, se debe principalmente a grandes frecuencias de estudiantes que obtienen una misma calificación. Esto sucede principalmente en las asignaturas que son consideradas de alta complejidad ya que suelen tener picos en la nota mínima necesaria para aprobar la asignatura. Por esta razón, puede existir un sesgo entre el valor asociado a la probabilidad de que la variable aleatoria que describe la calificación de esa asignatura esté por encima de un valor determinado y la proporción real de las calificaciones que superan este valor. De cualquier forma se presume que esa diferencia no afecta significativamente los análisis realizados en la metodología.

Debido a que la distribución Beta está definida únicamente en el intervalo $(0,1]$, es necesario para aplicar esta función, eliminar los registros con calificaciones iguales a cero, asumiendo que estos valores no represetan el comportamiento normal de los estudiantes, además el conjunto de registros restantes debe ser dividido por el factor 5.05 para asegurar la aplicación de esta distribución.

Diagrama de Flujo:



Ejemplo:

//MatrizFProbabilidadMaterias <- MatrizProMa(BaseDatos)
//showData(na.omit(as.data.frame(MatrizFProbabilidadMaterias[1])))

Visualización Salida:

	ID	N	DISTR	P1	P2	P.Ks
132 Constitución y Derecho Público	16153	158	Beta	25.61	2.94	0.11
137 Control-Mejoramiento del Suelo	4006	61	Beta	19.72	5.8	0.22
138 Control del Estrés	2817	18	Discreta	0.92	0.05	ND
144 Creatividad Org	15814	75	Beta	11.43	1.1	0.04
149 Cultivos de Uso Ilícito	19733	20	Discreta	0.88	0.05	ND
166 Desarrollo de la Inteligencia	2775	20	Beta	14.52	2.68	0.2
169 Desastres en Ingeniería	22721	335	Beta	4.37	1.16	0
177 Dinámica Estructural	4007	43	Beta	11.19	2.71	0.16
182 Diseño de Fundaciones	4008	225	Beta	12.91	4.47	0.13
183 Diseño en Concreto	3183	213	Beta	10.46	2.7	0.09
195 Ecología para todos	21845	19	Beta	13.57	2.76	0.75
202 Ecuaciones Diferenciales	1300	417	Beta	6.69	4	0

9.5.2.2 Función: Gráfico de histograma y FDP estimada

Apartado: Distribuciones de probabilidad.

Función: Gráfico de histograma y función de probabilidad estimada.

Hipervínculo: Código – Invocación

Sintaxis: GraficarProbHist(IDM)

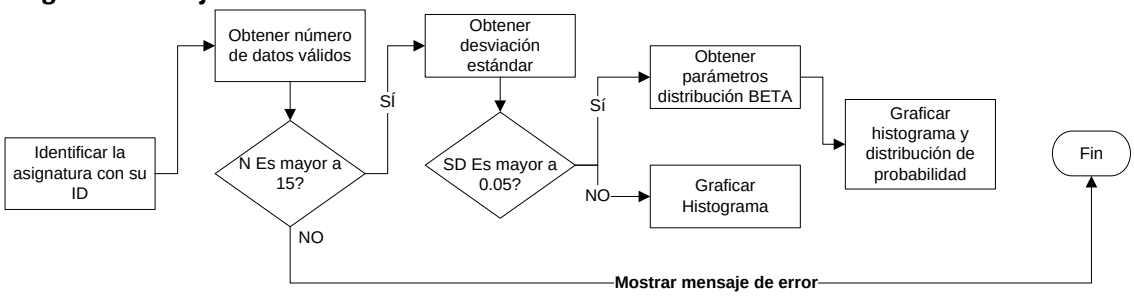
Parámetros: ID asignatura.

Salida: Histograma de la asignatura identificada con el ID y el gráfico de la curva de la distribución de probabilidad estimada.

Objetivo: Evidenciar la tendencia central, forma y dispersión de los datos, lo que permite tener una idea general del comportamiento de las calificaciones de una asignatura. De igual manera sirve para comparar gráficamente qué tan bien el conjunto de calificaciones se ajusta a la distribución de probabilidad.

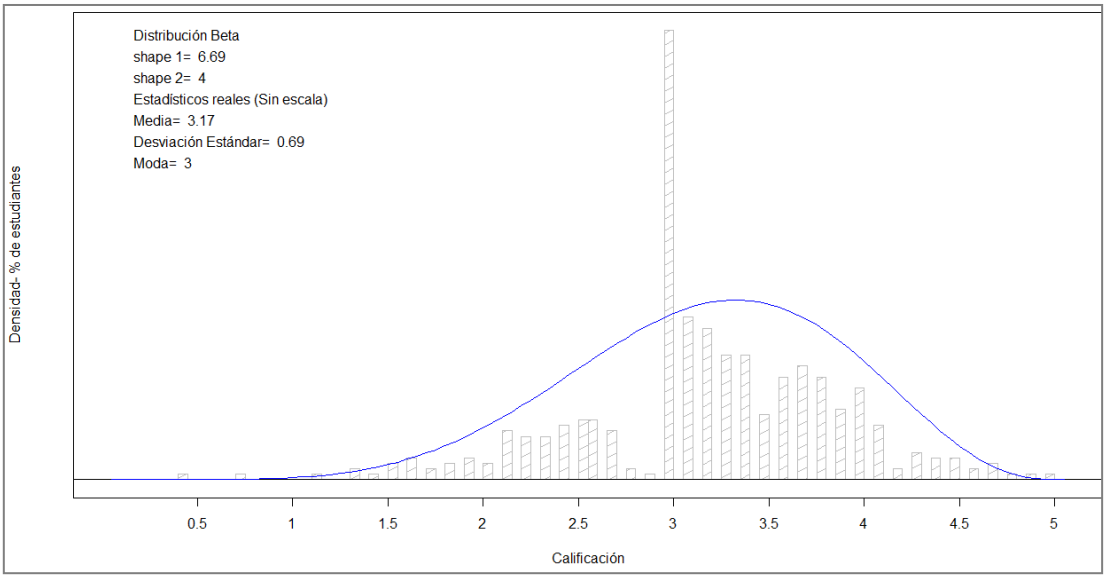
Precondiciones: Las distribuciones de probabilidad estiman adecuadamente el comportamiento de las calificaciones de la asignatura.

Diagrama de Flujo:



Ejemplo:
//GraficarProbHist(001300)

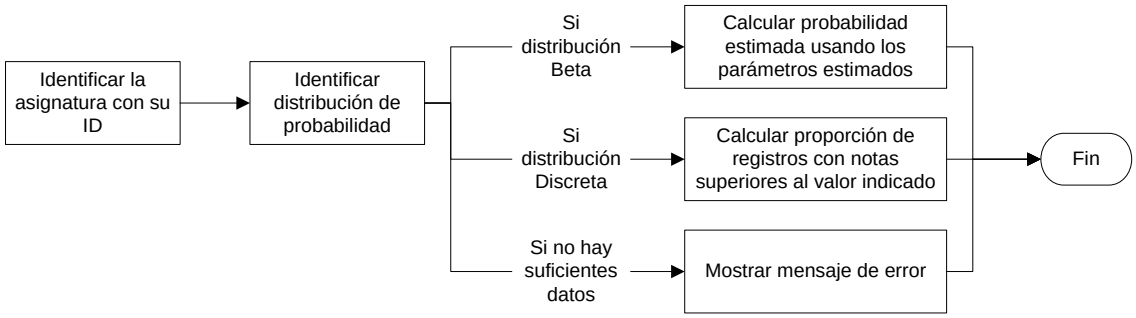
Visualización salida:



9.5.2.3 Función: Cálculo de probabilidades estimadas

Apartado: Distribuciones de probabilidad.
Función: Cálculo de probabilidades estimadas.
Hipervínculo: Código – Invocación
Sintaxis: Probestimada(nota, IDM)
Parámetros: Calificación, ID asignatura.
Función requerida: Matriz de Probabilidad por Asignatura.
Salida: Probabilidad de que un estudiante cualquiera obtenga en la asignatura identificada con el ID una calificación superior a la indicada. En particular si $\boldsymbol{x}$ es la variable aleatoria que representa las calificaciones de la asignatura, entonces el resultado de esta función será: $\boldsymbol{P(X \geq nota)}$, de igual manera es posible obtener el complemento del resultado, es decir $\boldsymbol{P(X < nota)}$ restando de una unidad el valor resultante de la función.
Objetivo: Calcular la probabilidad de obtener una calificación mayor a un valor estipulado, respondiendo algunas preguntas como, ¿Aceptando las precondiciones, cuál es la probabilidad de aprobar la asignatura?
Precondiciones: Los cálculos de estas probabilidades no tienen en cuenta ninguna característica intrínseca al estudiante. Además se presume que las distribuciones de probabilidad estimadas representan fielmente el comportamiento de las calificaciones de las asignaturas. De igual manera también se asume que los diferentes números de clases que representan (diferentes cursos, horarios, profesores) no tienen incidencia significativa en la calificación [56], [57].

Diagrama de Flujo:



Ejemplo:

//Probestimada(3, 001300)
//showData(MostrarProbabilidadResultante)

Visualización salida:

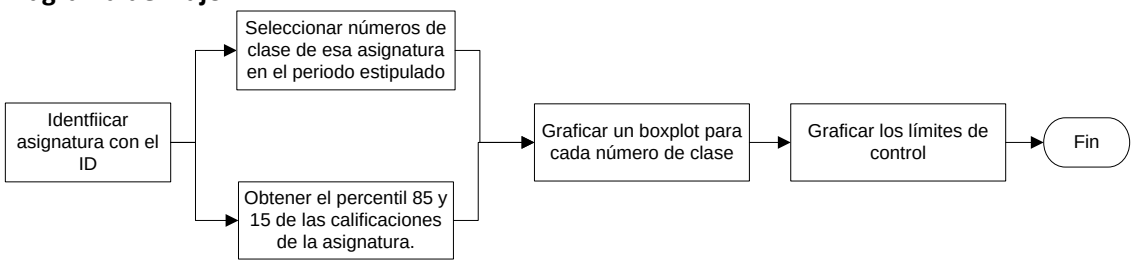
```
La probabilidad de sacar una nota mayor o igual a 3 en Ecuaciones Diferenciales
                                                                        60.3%
```

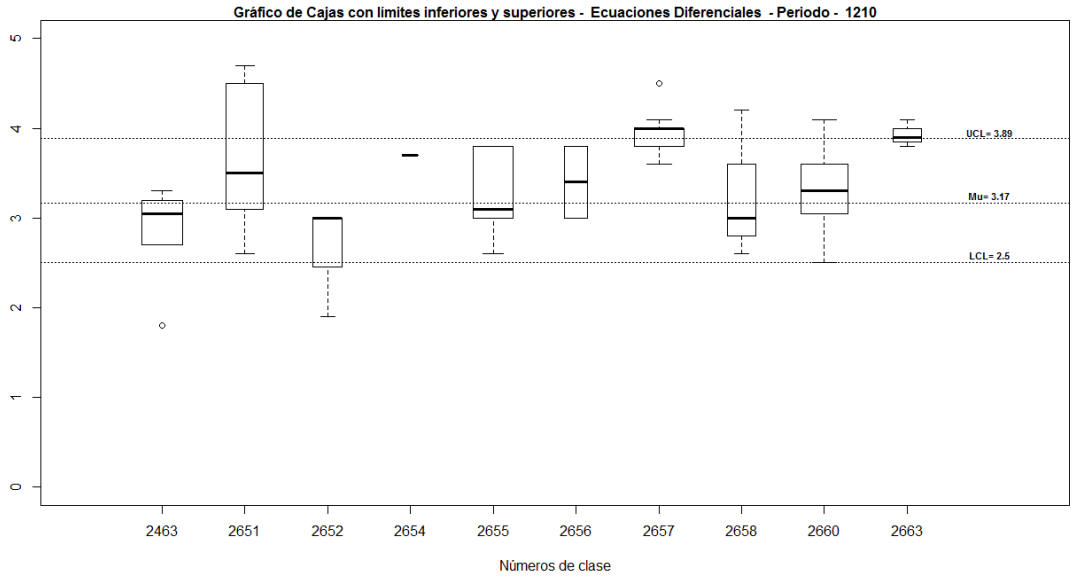
9.5.3 Apartado gráficos de control

Apartado estipulado para el seguimiento y control de las asignaturas y números de clase.

9.5.3.1 Función: Gráfico de cajas de control por asignatura y periodo

Apartado: Gráficos de control.
Función: Gráfico de cajas de control por asignatura y periodo.
Hipervínculo: Código – Invocación
Sintaxis: Grafcontrol(IDM, Periodo)
Parámetros: ID asignatura, Periodo académico.
Salida: Gráfico de cajas de cada número de clase que componen una asignatura en un periodo académico determinado. La plantilla del gráfico presenta los límites de control en el percentil no.15 y no. 85 del total de calificaciones disponibles de la asignatura en la base de datos.
Objetivo: Permitir la identificación de aquellos números de clase que están por fuera de los límites normales del comportamiento de la asignatura. Además abre una nueva oportunidad de investigación, usar la metodología propuesta como estrategia de seguimiento del resultado de la cátedra de los docentes.
Precondiciones: Los datos contenidos dentro del percentil no. 15 y no. 85 tienen un comportamiento corriente. Además, se presume que el número de datos válidos que contiene cada número de clase resulta suficiente para caracterizar su comportamiento.
Diagrama de Flujo:

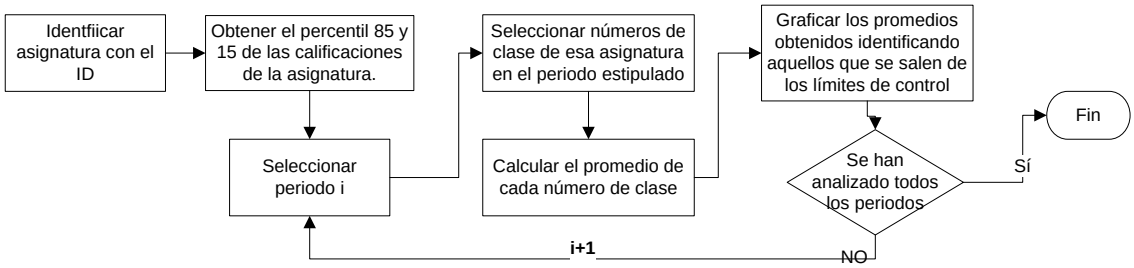


Ejemplo: *//Grafcontrol(001300, 1210)*
Visualización salida: 

9.5.3.2 Función: Gráfico de control por asignatura y números de clase

Apartado: Gráficos de control.
Función: Gráfico de control por asignatura y números de clase.
Hipervínculo: Código – Invocación
Sintaxis: Grafcontrolperiodoclases(IDM)
Parámetros: ID asignatura
Salida: Gráfico de control con la línea de tendencia en el tiempo del promedio general de la asignatura. Los límites de control están representados por el percentil no. 15 y no. 85 del total de calificaciones de la asignatura en la base de datos. Para cada periodo se indica la posición del promedio de la calificación obtenida en cada uno de los números de clase, indicando en color rojo aquellos que estén fuera de los límites de control. De igual manera se provee la tabla detallada de la información utilizada para la construcción del gráfico.
Objetivo: Identificar el número de clases que están por fuera de los límites de control por cada periodo de la asignatura estipulada, presentar la tendencia del promedio general de la asignatura y de la cantidad de número de clases fuera de control. Usando la tabla de información detallada es posible aislar los casos atípicos para su posterior seguimiento y control.
Precondiciones: Los datos contenidos dentro del percentil no.15 y no. 85 tienen un comportamiento corriente. Además se presume que el número de datos válidos que contiene cada clase resulta suficiente para caracterizar su comportamiento.

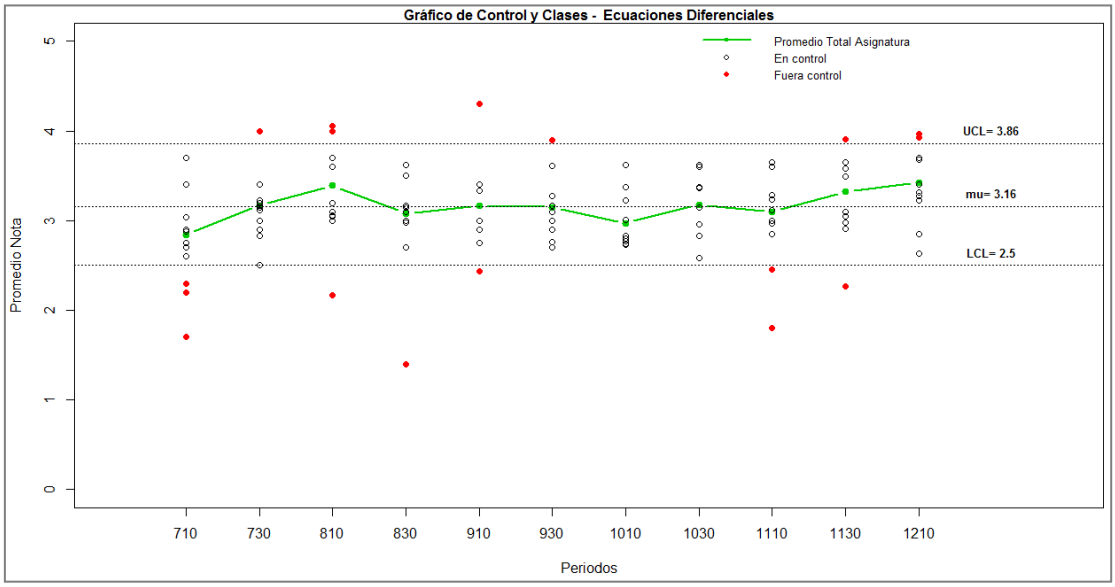
Diagrama de Flujo:



Ejemplo:

//DatosMateriaClases <- Grafcontrolperiodoclases(001300)
// showData(DatosMateriaClases)

Visualización salida:



	Nclase	Número de Datos	Periodo Académico	Promedio	Estado
1	2234	2	710	1.7	Fuera de Control
2	2249	4	710	2.88	Bajo Control
3	2254	5	710	3.04	Bajo Control
4	2259	3	710	2.7	Bajo Control
5	2275	1	710	3.4	Bajo Control
6	2281	1	710	2.3	Fuera de Control
7	2290	3	710	3.7	Bajo Control
8	2298	4	710	2.9	Bajo Control
9	2308	1	710	2.2	Fuera de Control
10	2341	2	710	2.75	Bajo Control
11	2370	2	710	2.6	Bajo Control
12	2835	1	730	3	Bajo Control
13	2843	1	730	3.2	Bajo Control
14	2861	1	730	3	Bajo Control
15	2914	9	730	3.4	Bajo Control

9.5.3.3 Función: Ranking de asignaturas con números de clase fuera de control

Apartado: Gráficos de control.
Función: Ranking de asignaturas con números de clase fuera de control.
Hipervínculo: Código – Invocación

Sintaxis: Variablesfueracontrol(Periodo)

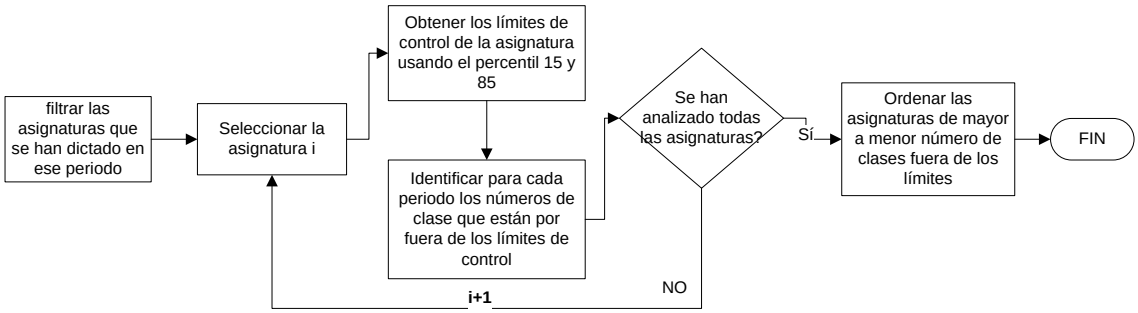
Parámetros: Periodo Académico

Salida: Tabla que presenta las asignaturas ordenadas según el número de clases fuera de control que tienen en un periodo determinado. También indica un estadístico de desviación, el cual se refiere al promedio del valor absoluto de cuánto sobrepasan los límites de control los números de clase que los exceden.

Objetivo: Enfocar los análisis a las asignaturas que tiene un mayor número de clases por fuera de los límites de control.

Precondiciones: Los datos contenidos dentro del percentil no. 15 y no. 85 tienen un comportamiento corriente, el número de datos válidos que tiene cada asignatura resulta significativo para realizar el análisis y caracterización del comportamiento de cada número de clase.

Diagrama de Flujo:



Ejemplo:

//MatrizVariablesFueraControl <- Variablesfueracontrol(1110)
//showData(MatrizVariablesFueraControl)

Visualización del salida:

ID	Asignatura	Número de Datos	Desviación	NExcendentes
2475	Vida y Horizontes creativos	26	0.5225	5
1342	Introducción a la Física	32	0.7988	4
1290	Álgebra Lineal	52	0.33	3
2253	Baloncesto, Iniciación.	8	0.145	3
2327	Cardio, acondicionamiento.	9	18.675	3
1340	Física Mecánica	69	0.776	3
2328	Fuerza, acondicionamiento.	7	1.4311	3
25104	Vínculos de Pareja	4	2.9072	3
1297	Cálculo Integral	56	0.065	2
1300	Ecuaciones Diferenciales	49	0.4925	2
1341	Fluidos y Termodinámica	58	0.0031	2
19589	Hoja de cálculo nivel avanzado	8	1.22	2
19588	Hoja de cálculo nivel básico	19	0.005	2
7033	Inglés 4	11	0.5	2
143	Introducción al Diseño	3	0.0045	2
4206	Pensamiento Algorítmico	63	0.17	2
3165	Squash, Iniciación.	7	0.3844	2
4000	Abastecimiento-Potabilización	3	0.0011	1
21657	Agro ecología en los andes	1	0.5184	1

9.5.4 Apartado análisis de agrupación

Apartado dirigido a la agrupación de asignaturas según algunas características comunes con el fin de crear perfiles de caracterización y poder en un futuro estandarizar su proceso de análisis y seguimiento.

9.5.4.1 Función: Agrupación de asignaturas electivas y complementarias

Apartado: Análisis de Agrupación.

Función: Agrupación de asignaturas electivas y complementarias.

Hipervínculo: Código – Invocación

Sintaxis: Grupoelecp (BaseDatos)

Parámetros: Base de datos académica depurada.

Salida: Cada uno de los registros de las asignaturas complementarias y electivas es agrupado según la naturaleza temática y la organización académica que imparta la asignatura.

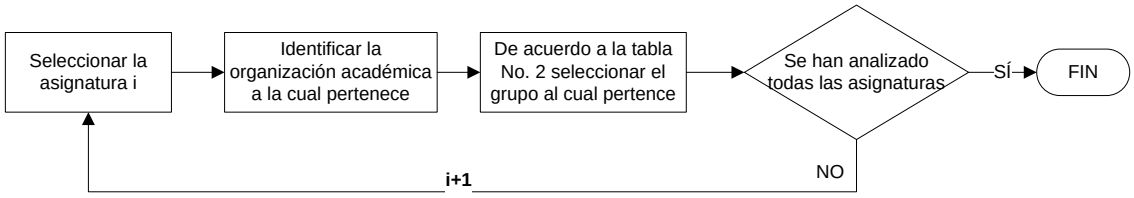
Tabla No. 6 Agrupación de asignaturas electivas (ELE) y complementarias (CP)

Grupo unificador	Organizaciones académicas que lo componen
Electivas Comunicación	DPT-COMUN
Electivas Música	DPT-MUSICA
Electivas Deportes	DPT-FORMACION
Electivas Derecho	DPT-PUBL, DPT-DLABOR, DPT-DFILDER, DPT-DPENAL.
Electivas Artes	DPT-ARESC, DPT-ARTVIS, DPT-LITER, DPT-ESTET, DPT-DESRUR
Electivas Salud	DPT-ESALUD, DPT-ECLIN
Electivas Lenguas	DPT-LENGS
Electivas Ingeniería Civil	DPT-ICIVIL, PRG-GVIALP
Electivas Arquidiseño	DPT-ARQUI, DPT-DISIND
Electivas Sicología	DPT-PSICOL
Electivas_Admonecono	DPT-ADMÓN, DPT-ECONOM
Electivas_Cienciasbasicas	DPT-PPROD, DPT-FÍSICA, DPT-ISIST, DPT-MATEM, DPT-ELECT
Electivas Sociales	DPT-TEOLOG,DPT-ANTROP, DPT-HGEOG, DPT-RELINT, DPT-FILOS, DPT-ETERR, DPT-CPOLÍT, DPT-SOCIOL.
Otro	Demás organizaciones académicas

Objetivo: Facilitar el análisis y entendimiento de los resultados de la metodología, enfocando los esfuerzos en los grupos relevantes de las asignaturas electivas y complementarias.

Precondiciones: Es válido agrupar asignaturas electivas y complementarias debido a que una misma asignatura puede tener doble naturaleza dependiendo de la elección del estudiante.

Diagrama de Flujo:



Ejemplo:

//AgrupacionELECP(BaseDatos)

Visualización salida: Esta función no presenta visualmente la agrupación únicamente crea un objeto que será utilizado por otras funciones de la metodología.

9.5.4.2 Función: Agrupación por K medias

Apartado: Análisis de Agrupación.

Función: Agrupación por K medias.

Hipervínculo: Código – Invocación

Sintaxis: Cluster(BaseDatos)

Parámetros: Base de datos académica depurada.

Función requerida: Agrupación de asignaturas electivas y complementarias.

Salida: Cada una de las asignaturas, incluyendo los grupos de asignaturas de la función anterior, se agrupará en dos conjuntos diferentes según los siguientes criterios:

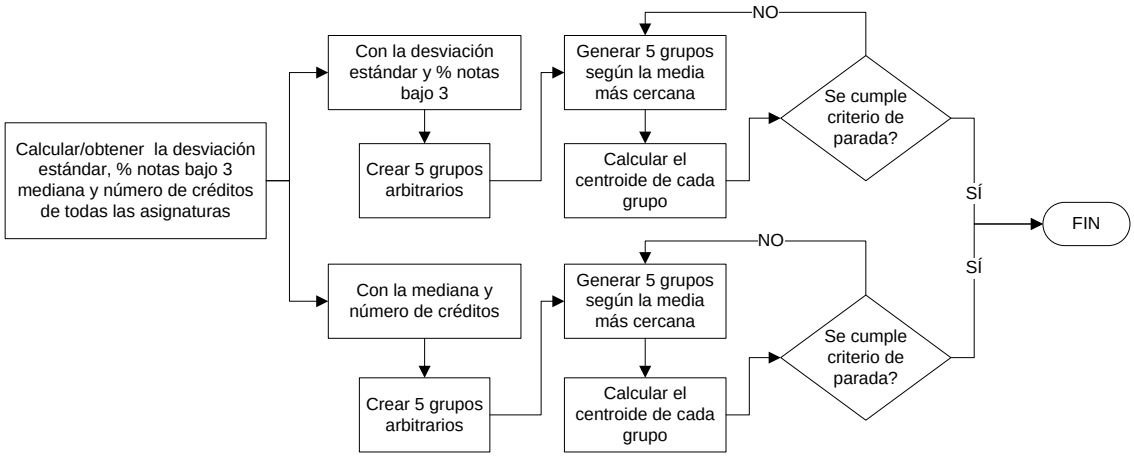
Tabla No. 7 criterios de agrupación a través de K medias.

Grupo	Variable discriminadora No. 1	Variable discriminadora No. 2
Grupo 1	Desviación estándar	% notas bajo 3
Grupo 2	Mediana	Número de créditos

Objetivo: Crear perfiles de asignaturas con las cuales se podría tener un procedimiento especializado y estandarizado de análisis.

Precondiciones: Cinco clases son suficientes y adecuadas para realizar la clasificación en cada uno de los dos grupos.

Diagrama de Flujo:



Ejemplo:

//DatosCluster <- Cluster()
//showData(DatosCluster[[4]])

Visualización del salida:

Nombre	C.Desv-%Perd	C.Med-Ncred	Promedio	Mediana	Desviación	Porcentajeperdidas
Abastecimiento-Potabilización	4	2	4.19	4.20	0.24	0.00
Álgebra Lineal	3	3	3.35	3.40	0.79	0.23
Análisis Estructural	1	3	3.75	3.80	0.65	0.10
Análisis Numérico	1	3	3.66	3.75	0.70	0.08
Cálculo Diferencial	3	3	3.01	3.10	0.83	0.34
Cálculo Integral	2	3	3.14	3.20	0.73	0.26
Cálculo Vectorial	2	3	3.03	3.00	0.66	0.32
Constitución y Derecho Público	4	4	4.54	4.60	0.28	0.00
Control-Mejoramiento del Suelo	5	2	3.92	4.00	0.43	0.00
Dinámica Estructural	5	2	4.06	4.10	0.43	0.00
Diseño de Fundaciones	1	3	3.67	3.80	0.74	0.06
Diseño en Concreto	1	3	3.99	4.00	0.67	0.05
Drenaje-Tratamiento Aguas R	5	2	3.99	4.00	0.57	0.07
Ecuaciones Diferenciales	2	3	3.16	3.10	0.73	0.23

9.5.4.3 Función: Gráfico de agrupación por K medias

Apartado: Análisis de Agrupación.

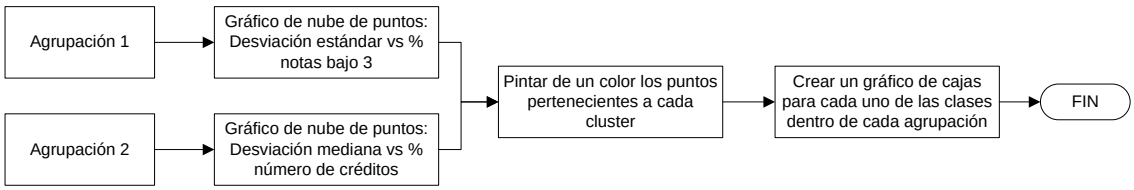
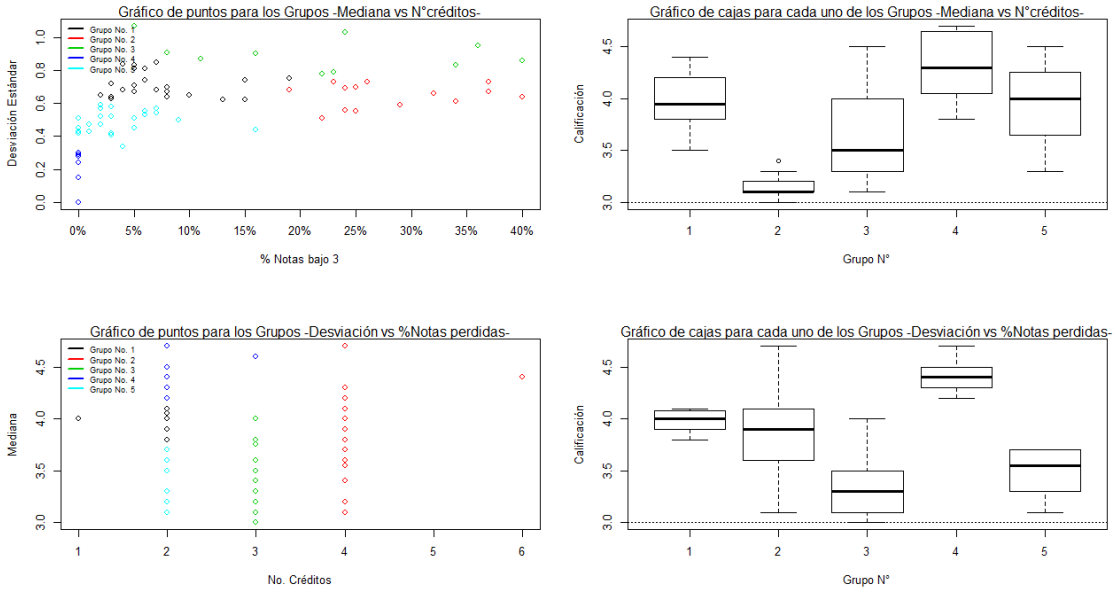
Función: Gráfico agrupación por K medias.

Hipervínculo: Código – Invocación

Sintaxis: Cluster(BaseDatos)

Parámetros: Base de datos académica depurada.

Salida: Gráfico de nube de puntos que presenta por color cada clase de cada grupo y gráfico de

cajas de cada clase por cada agrupación.
Objetivo: caracterizar cada uno de las clases de cada agrupación creada.
Precondiciones: El método de las K medias generó una agrupación lógica y pertinente.
Diagrama de Flujo:    **Ejemplo:** //DatosCluster <- Cluster()
Visualización salida: 

9.5.5 Apartado asignaturas re-cursasdas

Apartado enfocado a la descripción y análisis de las asignaturas que los estudiantes reprueban y re-cursan al menos una vez.

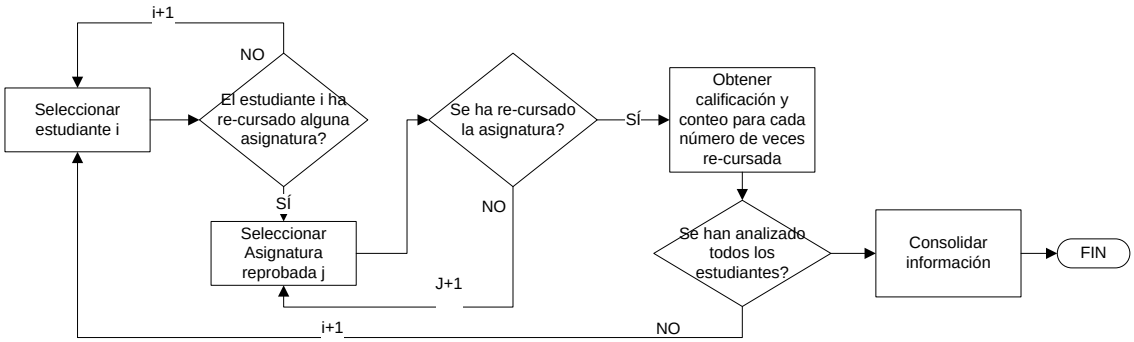
9.5.5.1 Función: Información de asignaturas re-cursadas

Apartado: Asignaturas re-cursadas.
Función: Información de asignaturas re-cursadas
Hipervínculo: Código – Invocación
Sintaxis: DatosAsignaturasRepetidas(BaseDatos)
Parámetros: Base de datos académica depurada.
Salida: Tabla que consolida el número de veces que se re-cursa una asignatura y para cada una de ellas indica el promedio de las calificaciones y el número de estudiantes que la re-cursaron.
Objetivo: Filtrar las asignaturas en las cuales el número de estudiantes que la re-cursan va en aumento o no está dentro de las condiciones esperadas.
Precondiciones: Los estudiantes que reprueban la asignatura por primera vez y no la vuelven a cursar no están dentro del análisis. Esto se debe principalmente a que si un estudiante pierde

por primera vez una asignatura tiene dos opciones: Retirarse del programa académico o re-cursar la asignatura. El grupo de estudiantes que elige la primera opción quizá está afectado por diferentes factores externos (Económicos, de convicción, etc.) lo cual se presume puede afectar significativamente las conclusiones de la aplicación de este apartado de la metodología.

Es probable que en algunas asignaturas el promedio de calificaciones disminuya si aumenta el número de veces cursada. Esto principalmente se debe a que este es un valor condicional que depende rotundamente de la calidad de los estudiantes en curso, por lo que se podría asumir que los estudiantes que re-cursan la asignatura varias veces no tienen la misma probabilidad de obtener una buena calificación que aquellos estudiantes que nunca han reprobado la asignatura.

Diagrama de Flujo:



Ejemplo:

//InformacionAsignaturasRepetidas <- DatosAsignaturasRepetidas()
//showData(InformacionAsignaturasRepetidas[[3]])

Visualización del salida:

ID	Asignatura	Número_de_veces_cursada	Promedio	Número_Datos
1299	Cálculo Vectorial	1	2.56	94
1299	Cálculo Vectorial	2	3.13	63
1299	Cálculo Vectorial	3	3.31	14
1299	Cálculo Vectorial	4	3.25	2
2327	Cardio, acondicionamiento.	1	2.5	2
2327	Cardio, acondicionamiento.	2	4.05	2
3155	Ciclismo Bajo Techo,principios	1	3.6	2
16153	Constitución y Derecho Público	1	4.73	3
2817	Control del Estrés	1	4.7	1
22721	Desastres en Ingeniería	1	2.75	17
22721	Desastres en Ingeniería	2	3.37	7
4008	Diseño de Fundaciones	1	1.89	10
4008	Diseño de Fundaciones	2	3.37	7
3183	Diseño en Concreto	1	2.85	11
3183	Diseño en Concreto	2	3.76	8
1300	Ecuaciones Diferenciales	1	2.51	83
1300	Ecuaciones Diferenciales	2	3.16	62
1300	Ecuaciones Diferenciales	3	3.15	6

9.5.5.2 Función: Gráfico FDP y tendencia de asignaturas re-cursadas

Apartado: Asignaturas re-cursadas.
Función: Gráfico FDP y tendencia de asignaturas re-cursadas.
Hipervínculo: Código – Invocación
Sintaxis: Graficoperdidasporasignatura(InformacionAsignaturasRepetidas, IDM)
Parámetros: Tabla de información resultante de usar la función Información de asignaturas re-cursadas, ID asignatura.

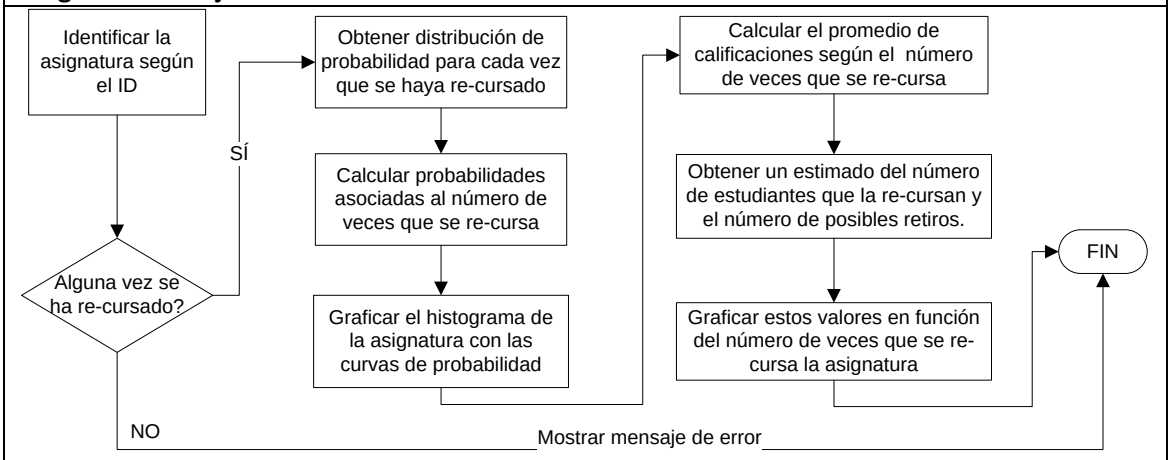
Función necesaria: Información asignaturas re-cursadas

Salida: Histograma de las calificaciones de la asignatura estipulada, indicando para cada una de las veces que se re-cursa la distribución de probabilidad asociada, gráfico que presenta la tendencia del promedio y número de estudiantes que cursan la asignatura según el número de veces re-cursada, tabla con las probabilidades de obtener una nota mayor a 2.0, 2.5, 3.0 3.5, 4.0 y 4.5 para cada una de las veces que se re-cursa la asignatura.

Objetivo: Identificar los cambios en las tendencias de las calificaciones según el número de veces que una asignatura es re-cursada, mostrar el cambio en el promedio y número de estudiantes que cursan una asignatura un número determinado de veces, analizar si los cambios en las curvas de probabilidad y las tendencias sugieren una relación entre el número de veces que un estudiante cursa una asignatura y la probabilidad de aprobarla.

Precondiciones: Los números de datos de cada uno de los subconjuntos de registros creados según el número de veces que se re-cursa una asignatura son suficientes para estimar y graficar la función de probabilidad asociada. Para el cálculo de probabilidades se requiere al menos 15 datos válidos para estimar su valor a través de una distribución de probabilidad, de lo contrario se usará la proporción de casos favorables.

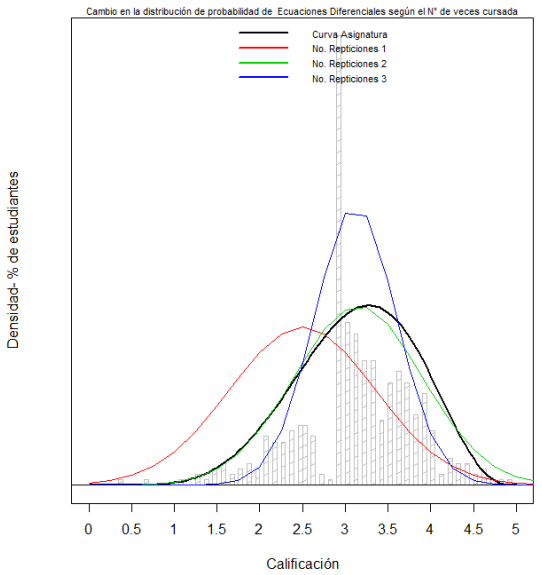
Diagrama de Flujo:

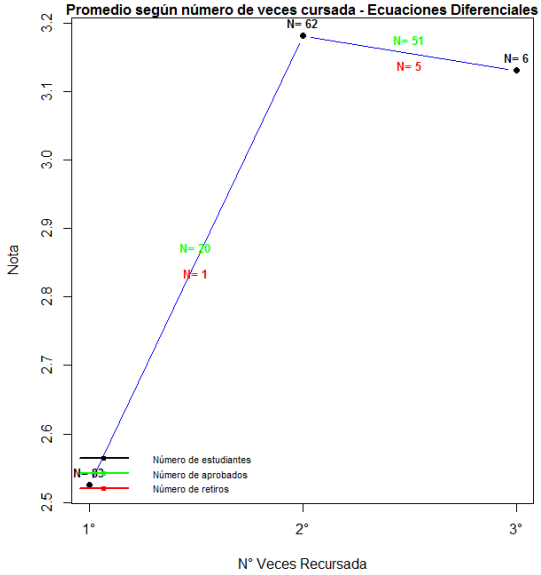


Ejemplo:

//Graficoperdidasporasignatura(InformacionAsignaturasRepetidas[[1]],1300)

Visualización del salida:





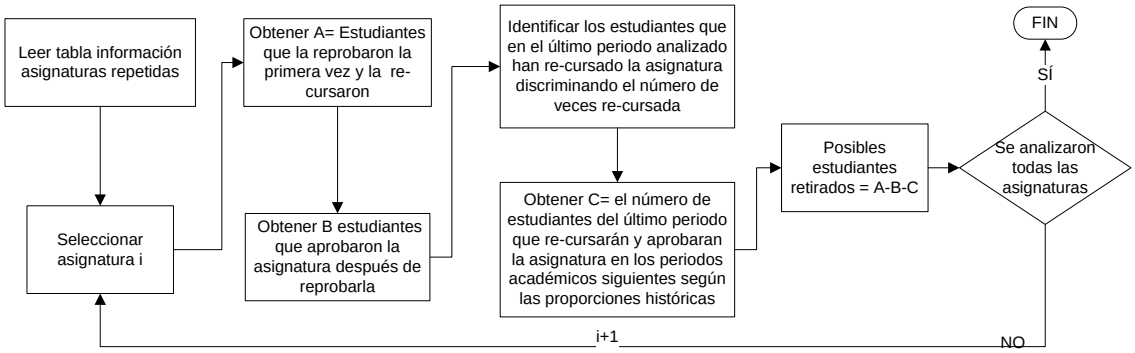
Nota	P-Asignatura	Asignatura N.Rep -1	Asignatura N.Rep -2	Asignatura N.Rep -3
2	93.75%	72.96%	94.06%	100%
2.5	81.46%	51.16%	81.59%	100%
3	60.39%	29%	59.47%	83.33%
3.5	34.41%	12.8%	33.71%	16.67%
4	12.31%	4.29%	14%	16.67%
4.5	1.53%	1.07%	4.09%	0%

9.5.5.3 Función: Posibles estudiantes retirados del programa académico

Apartado: Asignaturas re-cursadas.
Función: Posibles estudiantes retirados del programa académico.
Hipervínculo: Código – Invocación
Sintaxis: NumeroRetiros(InformacionAsignaturasRepetidas)
Parámetros: Tabla de información resultante de usar la función Información de asignaturas re-cursadas.
Función necesaria: Información asignaturas re-cursadas
Salida: Tabla que indica para cada asignatura el número de estudiantes que la re-cursan por primera vez y el número de estudiantes que aún no ha aprobado o re-cursado de nuevo.
Objetivo: Identificar las asignaturas que posiblemente causan mayor deserción de estudiantes.
Precondiciones: No se tiene en cuenta la secuencia de cada estudiante en particular, sino el movimiento de todos los estudiantes que según la base de datos la han cursado un número determinado de veces. Debido al muestreo realizado sobre la base de datos general no se tiene en cuenta parte de la trayectoria de asignaturas re-cursadas de los estudiantes con registros anteriores al primer periodo tomado como referencia para crear la base de datos. Las proporciones de estudiantes que re-cursan la asignatura después de haberla reprobado un número determinado de veces representa fielmente la probabilidad de que esto suceda. De igual manera se supone que los estudiantes re-cursan la asignatura en el periodo académico inmediatmente siguiente, por lo que el valor esperado de la proporción de estudiantes que aprueban la asignatura después de re-cursarla un número determinado de veces se aplica sólo

para el último periodo.

Diagrama de Flujo:



Ejemplo:

//Posiblesretiros <- NumeroRetiros(InformacionAsignaturasRepetidas[[1]])
//showData(Posiblesretiros)

Visualización salida:

	ID	N° reprobados 1er vez	Número de posibles retiros	%
Cálculo Diferencial	1295	98	12	12%
Cálculo Integral	1297	83	13	16%
Cálculo Vectorial	1299	94	8	9%
Cardio, acondicionamiento.	2327	2	0	0%
Ciclismo Bajo Techo,principios	3155	2	0	0%
Constitución y Derecho Público	16153	3	0	0%
Control del Estrés	2817	1	0	0%
Desastres en Ingeniería	22721	17	3	18%
Diseño de Fundaciones	4008	10	0	0%
Diseño en Concreto	3183	11	0	0%
Ecuaciones Diferenciales	1300	83	5	6%

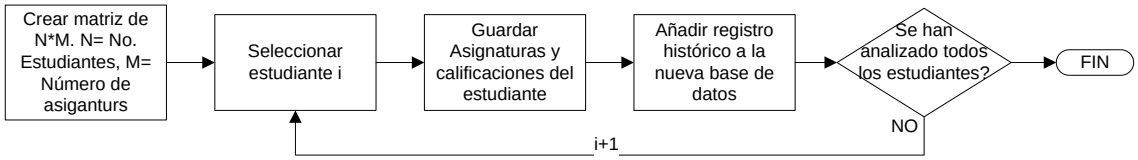
9.5.6 Apartado análisis de relaciones entre asignaturas

Apartado dirigido a identificar si algunas asignaturas tienen una influencia en las calificaciones de otras con el fin de poder crear estrategias de refuerzo desde etapas tempranas del proceso educativo.

9.5.6.1 Función: Base de datos orientada a estudiantes

Apartado: Análisis de relaciones entre asignaturas
Función: Base de datos orientada a estudiantes
Hipervínculo: Código – Invocación
Sintaxis: RegistroPorEstudiante(BaseDatos)
Parámetros: Base de datos académica depurada.
Función requerida: Agrupación de asignaturas electivas y complementarias.
Salida: Base de datos creada con el registro histórico de cada estudiante. Los estudiantes representan las filas y el total de asignaturas (con la agrupación de electivas y complementarias) representan las columnas, los datos contenidos en la base de datos son la calificación obtenida por ese estudiante en esa asignatura.
Objetivo: Facilitar el análisis de correlación entre asignaturas.
Precondiciones: Ninguno.

Diagrama de Flujo:



Ejemplo:

// BaseDatosPor estudinate <-RegistroPorEstudiante(BaseDatos)
// showData(BaseDatosPorestudiante)

Visualización Salida:

	Abastecimiento-Potabilización	Álgebra Lineal	Análisis Estructural	Análisis Numérico	Cálculo Diferencial
10091080	NA	4.2	4.1	3.4	3.3
10091133	NA	2.4	4.0	3.6	4.3
10091322	NA	3.6	NA	NA	3.3
10091333	NA	NA	NA	NA	4.4
10091589	NA	NA	NA	NA	1.7
10091732	NA	3.7	4.7	4.4	2.8
10091887	NA	4.1	3.8	4.3	3.4
10091957	NA	3.0	2.6	4.0	2.5
10092027	NA	2.3	NA	3.6	2.5
10092028	NA	3.7	3.8	3.9	1.9
10092178	NA	NA	3.2	3.7	NA
10092446	NA	2.9	4.2	3.4	3.0
10092462	NA	3.3	NA	NA	2.3
10092565	NA	3.0	4.1	5.0	3.0

9.5.6.2 Función: Matriz de correlaciones y asignaturas de mayor relación lineal

Apartado: Análisis de relaciones entre asignaturas.

Función: Matriz de correlaciones y asignaturas con mayor relación lineal.

Hipervínculo: Código – Invocación

Sintaxis: MatrizRelacionesPares(Base de datos orientada a estudiantes, % datos válidos)

Parámetros: Base de datos orientada a estudiantes y porcentaje de datos válidos mínimo necesario para que una asignatura entre en el análisis de correlaciones.

Función requerida: Base de datos orientada a estudiantes

Salida: Matriz NxN, donde N es el número de asignaturas, en la cual se indica la correlación de cada asignatura con las demás, expresada a través del coeficiente de correlación lineal de Pearson. Tabla que presenta las 10 asignaturas con mayor relación indicando el coeficiente de correlación de Pearson y la significancia de la siguiente prueba de hipótesis.

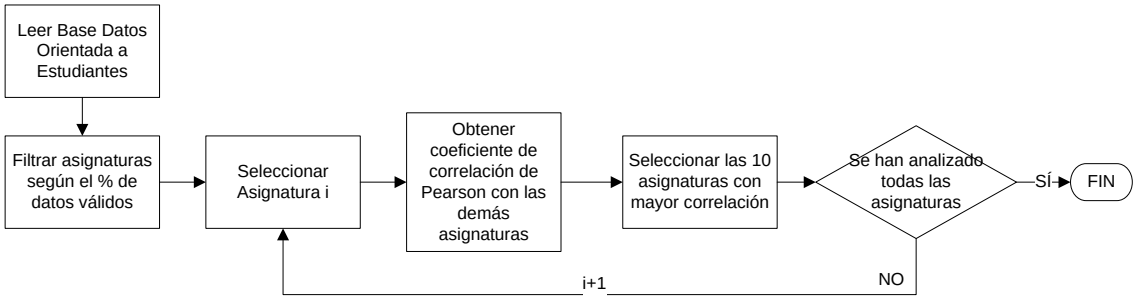
$$\text{Ho: } \rho = 0$$
$$\text{Ha: } \rho \neq 0$$

Donde ρ representa al coeficiente de correlación lineal de Pearson.

Objetivo: Presentar las asignaturas que tienen una mayor relación lineal con el fin de realimentar el proceso educativo y fortalecer el resultado académicos de las asignaturas dependientes.

Precondiciones: Sólo se tiene en cuenta la relación lineal existente entre cada pareja de asignaturas. Es preciso incluir un porcentaje de datos válidos en la función debido a que al momento de crear la matriz orientada a los registros históricos de cada estudiante, existirán muchos casos en los cuales no se haya cursado aún una asignatura, ya sea porque aún no se tiene las condiciones académicas o porque no hacen parte de los intereses de los estudiantes. Se recomienda usar un porcentaje de datos válidos del 20% para asegurar que el indicador de correlación de Pearson sea robusto.

Diagrama de Flujo:



Ejemplo:

//MatrizRelacionesPares(MatrizRegistroPorEstudiante, 20%)

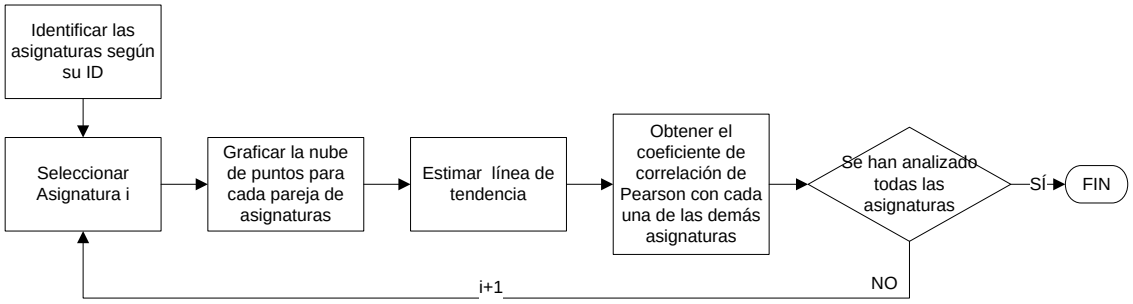
Visualización salida:

	1°		
Álgebra Lineal	Química de Materiales	R=0.5	P=0%
Análisis Estructural	Diseño en Concreto	R=0.43	P=0%
Análisis Numérico	Mecánica de Fluidos	R=0.43	P=0%
Cálculo Diferencial	Química de Materiales	R=0.71	P=0%
Cálculo Integral	Física Mecánica	R=0.44	P=0%
Cálculo Vectorial	Probabilidad y Estadística	R=0.41	P=0%
Diseño de Fundaciones	Diseño en Concreto	R=0.42	P=0%
Diseño en Concreto	Hidráulica Aplicada	R=0.49	P=0%
Ecuaciones Diferenciales	Fluidos y Termodinámica	R=0.38	P=0%
Epistemología de la Ingeniería	Significación Teológica	R=0.42	P=0%
Estática	Fluidos y Termodinámica	R=0.47	P=0%
Expresión Gráfica y Geometría	Química de Materiales	R=0.71	P=0%
Física Mecánica	Investigación de Operaciones	R=0.47	P=0%
Fluidos y Termodinámica	Estática	R=0.47	P=0%
Geología	Materiales de Construcción	R=0.35	P=0%
Hidráulica Aplicada	Hidrología	R=0.51	P=0%

9.5.6.3 Función: Gráfico de correlación de asignaturas

Apartado: Análisis de relaciones entre asignaturas
Función: Gráfico de correlación de asignaturas.
Hipervínculo: Código – Invocación
Sintaxis: Graficapuntosycorrelaciones(IDM1, IDM2, IDM3,…IDM10)
Parámetros: ID de asignaturas a graficar.
Salida: Gráfico matricial que presenta en la diagonal inferior una nube de puntos con su respectiva línea de tendencia estimada por cada pareja de asignaturas y en la diagonal superior presenta el coeficiente de correlación de Pearson con su tamaño de letra proporcional al valor del coeficiente.
Objetivo: Analizar de manera sencilla las relaciones lineales entre un grupo de asignaturas.
Precondiciones: Sólo se tiene en cuenta la relación lineal existente entre cada pareja de asignaturas.

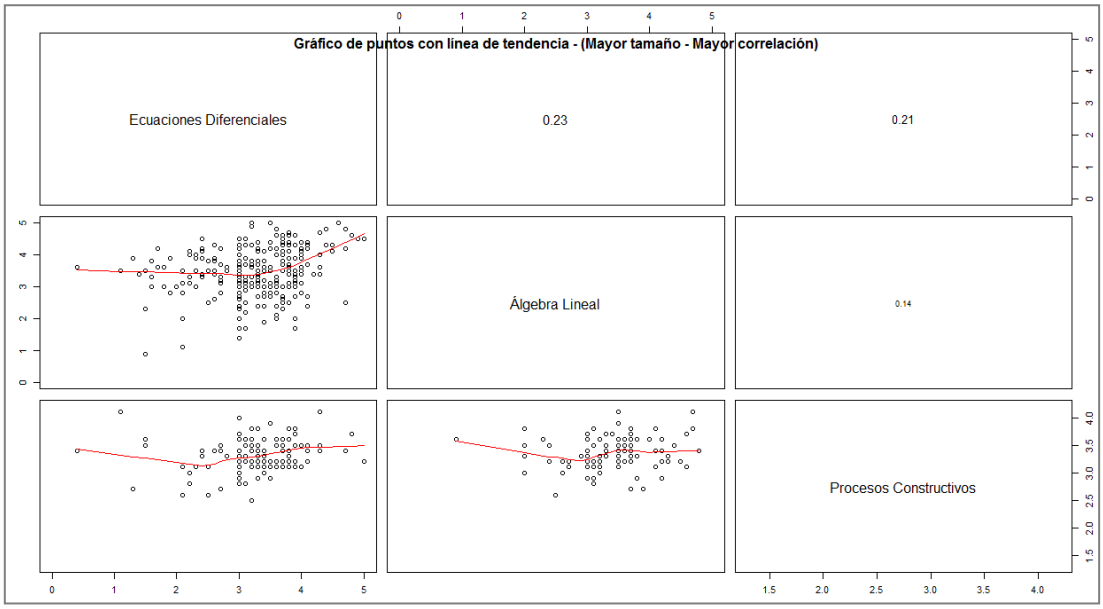
Diagrama de Flujo:



Ejemplo:

// Graficapuntosycorrelaciones(c(1300 ,1290, 4042))

Visualización salida:



9.5.7 Apartado Análisis de requisitos

Apartado dirigido al análisis de la matriz de pre-requisitos. Genera información para realizar el análisis de requisitos de una asignatura.

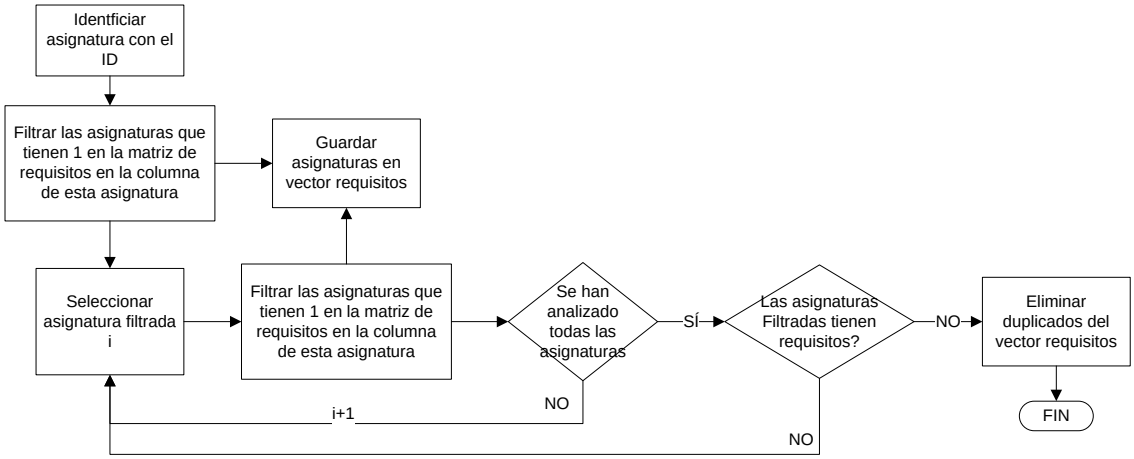
9.5.7.1 Función: Requisitos por asignatura

Apartado: Análisis de requisitos.
Función: Requisitos por asignatura.
Hipervínculo: Código – Invocación
Sintaxis: prerequisitos(IDM, MatrizRequisitosEspeciales, Matrizprecedencias)
Parámetros: ID asignatura, tabla de información de requisitos especiales, matriz de pre-requisitos.
Salida: Tabla que presenta cada una de las asignaturas que se deben aprobar para estar en condiciones académicas de inscribir la asignatura estipulada, indicando para cada una de ellas la secuencia de periodos en las que se podrían cursar de manera consecutiva.
Objetivo: Facilitar la revisión del cumplimiento de condiciones académicas para inscribir una

asignatura en particular, ofrecer estos conjuntos de asignaturas como entrada a modelos de tomas de decisiones.

Precondiciones: Ninguno.

Diagrama de Flujo:



Ejemplo:

// prerequisitos(001300)
//showData(Asignaturasrequisitos)

Visualización salida:

Asignaturas Requisitos de	Ecuaciones Diferenciales Periodos
Cálculo Integral	2
Álgebra Lineal	1
Cálculo Diferencial	1

9.5.7.2 Función: Selección de asignaturas candidatas

Apartado: Análisis de precedencias

Función: Selección de asignaturas candidatas.

Hipervínculo: Código – Invocación

Sintaxis: Asignaturaspuedecursar(DatosEstudianteRequisitosCumplidos, , MatrizRequisitosEspeciales, Matrizprecedencias)

Parámetros: Registro histórico del estudiante, tabla de información de requisitos especiales, matriz de pre-requisitos.

Salida: Tabla que presenta las asignaturas que el estudiante está en condiciones académicas de inscribir según el análisis de requisitos, número de créditos académicos aprobados y requisito de idioma extranjero.

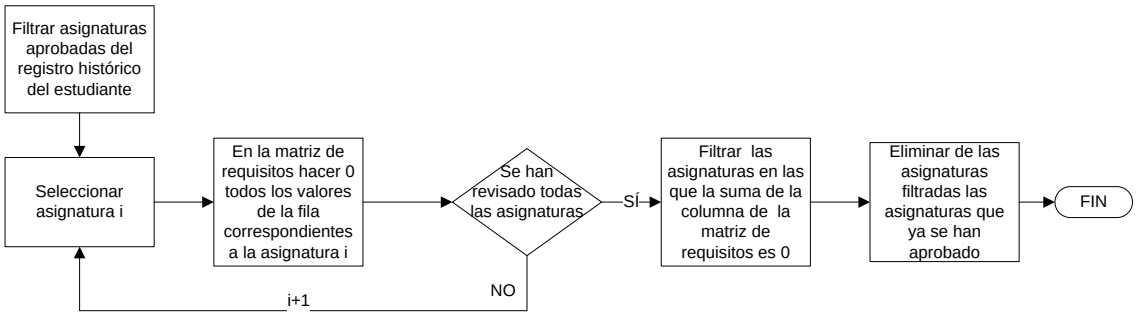
Dado un conjunto total de asignaturas pertenecientes al núcleo de formación fundamental o énfasis $\mathbf{T}$, un conjunto de asignaturas aprobadas $\mathbf{C}$ y una matriz de pre-requisitos $\mathbf{M}$. Es necesario encontrar el conjunto de asignaturas $\mathbf{A}$, $\mathbf{A} \in \{\mathbf{T} - \mathbf{C}\}$ que un estudiante está en capacidades académicas de cursar en el siguiente periodo académico. Para esto supongamos que se quiere conocer cuál es el conjunto de asignaturas que se puede cursar si el estudiante entra a primer semestre, es decir $\mathbf{C} = \emptyset$, según la definición de $\mathbf{M}$, si la suma de la columna jotaésima $\mathbf{M}.\mathbf{j} = \mathbf{0}$, entonces la asignatura con índice j no tiene requisitos y puede ser cursada en el primer semestre. Ahora, para cualquier conjunto de asignaturas aprobadas $\mathbf{C}$, se sabe

que cada una de las asignaturas $\mathbf{T} - \mathbf{C}$ representan una columna en $\mathbf{M}$ y que cada uno de sus componentes representa la dependencia hacia otras asignaturas, por lo que si hacemos que cada uno de los componentes fila $\mathbf{C}$ de la matriz $\mathbf{M}$, sea 0, las asignaturas candidatas $\mathbf{A}$, serían aquellas que cumplan $\mathbf{M}.\mathbf{j} = \mathbf{0} \ \forall \mathbf{j} \in \{\mathbf{T} - \mathbf{C}\}$.

Objetivo: Facilitar la revisión y selección del conjunto de asignaturas candidatas a inscribir según el registro académico histórico de un estudiante.

Precondiciones: Ninguno.

Diagrama de Flujo:



Ejemplo:

// Asignaturaspuedecursar(DatosEstudianteRequisitosCumplidos)
//showData(Asignaturascandidatasacursar)

Visualización salida:

```
Asignaturas disponibles para cursar
                    Cálculo Integral
                     Física Mecánica
               Pensamiento Algorítmico
          Topografía y Fotogrametría
          Materiales de Construcción
           Laboratorio de Materiales
```

9.5.8 Apartado valores esperados de calificaciones

Apartado dirigido a estimar el valor esperado de una asignatura en función de las calificaciones obtenidas a lo largo del programa académico. Definiendo $\boldsymbol{G}$ al conjunto de asignaturas vistas anteriormente, y a $\boldsymbol{x}$ la asignatura a analizar, se pretende identificar el subconjunto $\boldsymbol{G'} \in \boldsymbol{G}$, que tiene incidencia en los resultados de $\boldsymbol{x}$, ya sea de manera directa o a través de su interacción con otras variables del conjunto y estimar su aporte a la calificación esperada de la asignatura $\boldsymbol{x}$.

9.5.8.1 Función: Cálculo del valor esperado de las calificaciones de asignaturas candidatas

Apartado: Análisis de correlaciones.

Función: Cálculo del valor esperado de la calificación de asignaturas candidatas.

Hipervínculo: Código – Invocación

Sintaxis: EstimarNota(DatosEstudianteRequisitosCumplidos)

Parámetros: Registro histórico del estudiante.

Función requerida: Selección de asignaturas candidatas, Base de datos orientada a estudiantes.

Salida: Tabla que indica para cada una de las asignaturas candidatas el valor esperado de la

calificación según su registro histórico.

Este análisis se realiza a través de un modelo de regresión que tendrá en cuenta como variables predictoras de una asignatura en particular, todas aquellas asignaturas que el estudiante ya ha cursado y que tienen un coeficiente de correlación de Pearson significativamente diferente de 0, además se tiene en cuenta todas las interacciones por pares entre las asignaturas predictoras.

El modelo de regresión para una asignatura en particular es:

$$Y = \sum_{i=1}^{k} \beta_k x_k + \sum_{i=1}^{k} \sum_{\forall j \neq i} \beta_{ij} x_i * x_j + e$$

En donde el primer componente corresponde al aporte de cada una de las asignaturas predictoras, el segundo componente corresponde a las interacciones por pares de asignaturas y el tercer componente el error aleatorio.

Además, la metodología utiliza el proceso de regresión paso a paso (*stepwise regression*) hacia adelante para evitar que variables que comparten la misma información sobre la variable independiente aporten error al modelo. Por ejemplo, cálculo diferencial, cálculo integral y álgebra lineal podrían tener información similar acerca de la variable independiente ecuaciones diferenciales, por lo que agregar las 3 variables podría aportar más error que información útil al modelo.

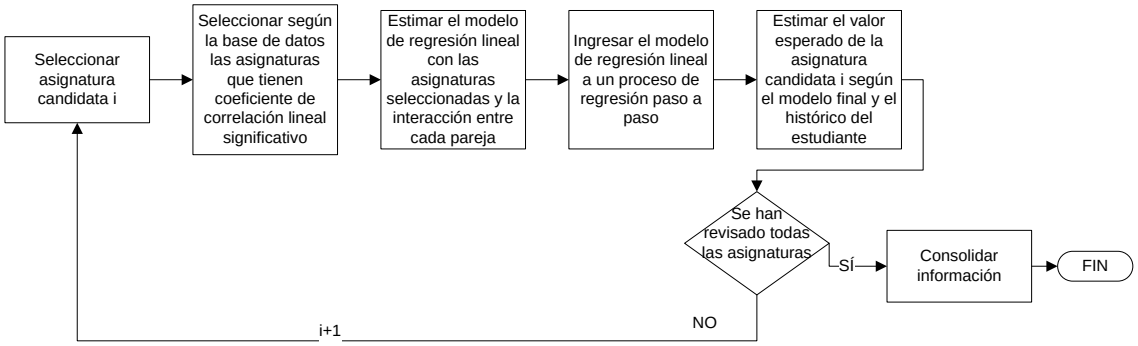
Para el modelo final de regresión se presenta el indicador R^2 ajustado, el cual expresa el porcentaje de la variabilidad de la asignatura dependiente explicada a través de las variables predictoras.

Las asignaturas ELE y CP van dirigidas a enriquecer los conocimientos interdisciplinarios de los estudiantes, por lo que se las debe tratar según la agrupación obtenida por la función 'Agrupacionelecp' para evitar sesgar las convicciones de aprendizaje de cada estudiante. Por esto, para las asignaturas candidatas de naturaleza CP se debe estimar su calificación según el promedio histórico de las asignaturas impartidas por ese programa académico en la base de datos académica. Por otra parte las asignaturas ELE representan por lo general actividades extracurriculares que requieren aptitudes específicas, por lo que se debe estimar su calificación según el promedio de calificaciones obtenidas por el estudiante en las asignaturas de esta naturaleza.

Objetivo: Permitir el cálculo del promedio ponderado acumulado esperado si el estudiante inscribe un conjunto de asignaturas en particular. Por otra parte el valor esperado de cada asignatura candidata le sirve al estudiante como una prealerta para identificar aquellas asignaturas en las que podría obtener una baja calificación si mantiene el ritmo de estudio actual.

Precondiciones: Sólo se tiene en cuenta la relación lineal y la interacción de cada pareja de asignaturas. Se asume que sólo aquellas asignaturas del registro histórico que tienen según la base de datos una correlación significativa al 5% con la asignatura candidata son las que deben incluirse dentro del modelo de regresión paso a paso. La función asumirá que todas las asignaturas impartidas por la organización académica que el estudiante indica como complementaria son en efecto complementarias, mientras que las demás asignaturas que no pertenecen al énfasis ni al núcleo fundamental serán etiquetadas como ELE.

Diagrama de Flujo:



Ejemplo:

//NotasEstimadas <- EstimarNota(DatosEstudianteRequisitosCumplidos)
//showData(NotasEstimadas)

Visualización salida:

	Asignatura	Nota Esperada	Naturaleza	R2 Ajus	ID
1	Cálculo Integral	3.4	NFFyENF	0.18	1297
2	Física Mecánica	3.4	NFFyENF	0.31	1340
3	Pensamiento Algorítmico	3.8	NFFyENF	0.43	4206
4	Topografía y Fotogrametría	3.8	NFFyENF	0.39	4180
5	Materiales de Construcción	4	NFFyENF	0.33	4033
6	Laboratorio de Materiales	3.7	NFFyENF	0.15	4030
7	Electivas_Deportes	5	ELE	<NA>	<NA>
8	Complementaria DPT-LENGS	3.9	CP	<NA>	<NA>

9.6 Modelo matemático de selección de asignaturas

De acuerdo al análisis de los modelos propuestos en la revisión bibliográfica para las diversas variaciones del problema KN y según contextualización del proceso de selección de asignaturas, a continuación se presenta el modelo matemático que describe el proceso de selección de asignaturas como un problema de programación matemática.

$$\text{Max} \ \frac{P_a * C_c + \sum_{\forall_i} c_i E[w_i|G]\, x_i}{C_c + \sum_{\forall_i} c_i x_i} \qquad (1)$$

$$\text{S. A.}$$

$$C_{min} \leq \sum_{\forall_i} c_i x_i \leq C_{max} \qquad (2)$$

$$Cele_{min} \leq \sum_{\forall_i} c_i x_i \leq Cele_{max} \qquad \forall\, x_i \in \{ELE\}\ (3)$$

$$Ccp_{min} \leq \sum_{\forall_i} c_i x_i \leq Ccp_{max} \qquad \forall\, x_i \in \{CP\}\ (4)$$

$$x_i \in \{0,1\} \qquad (5)$$

(1) corresponde a la función objetivo, que en este caso está representada por el promedio académico ponderado acumulado esperado de un estudiante según sus datos históricos y según el conjunto de asignaturas que se inscriban en el siguiente periodo académico, en donde P_a y C_c corresponden al promedio acumulado y al número de créditos acumulado que el estudiante lleva hasta el momento. $E[w_i|G]$ Representa el valor esperado de la nota de la asignatura i según un conjunto de asignaturas cursadas G, mientras que c_i y x_i son el número de créditos de la asignatura i y la variable binaria de decisión que indica si la asignatura i se inscribe en el siguiente semestre o no.

Esta función objetivo está restringida al número de créditos totales que es posible inscribir en un periodo académico según la modalidad de matrícula inscrita y el número mínimo de créditos que se está dispuestos a inscribir (2), y las restricciones asociadas al número máximo y mínimo de créditos destinados a asignaturas electivas y complementarias que se deseen cursar (3) ,(4).

Para evitar que esta metodología sesgue la convicción de aprendizaje de los estudiantes, las asignaturas electivas y complementarias se deben manejar de manera general, sin sugerir alguna asignatura en particular que podría desviar y viciar los objetivos propios del estudiante. Por lo tanto los grupos ELE y CP contarán cada uno con un número de asignaturas igual al número máximo permitido para cada naturaleza y tendrán el mismo valor esperado de la calificación, además de 1 solo crédito académico. Esta medida es necesaria para poder formular el proceso de decisión adecuadamente y recomendar el número de créditos de asignaturas electivas o complementarias que el estudiante debería inscribir.

Debido a que este modelo tiene varias restricciones y además su función objetivo no es lineal, podemos asumir al problema de selección de asignaturas como un problema de la mochila no lineal con restricciones múltiples (NLMKKN).

9.6.1 Modelo lineal relajado de selección de asignaturas

Para facilitar un posible método de solución se presenta la relajación lineal del modelo anterior, en el cual su función objetivo es remplazada por la suma ponderada de las diferencias de la nota esperada de cada asignatura con el promedio ponderado acumulado del estudiante a la fecha.

$$\text{Max} \sum_{\forall_i} c_i(E[w_i|G] - P_a)\, x_i \quad (1)$$

$$\text{S. A.}$$

$$C_{min} \leq \sum_{\forall_i} c_i x_i \leq C_{max} \quad (2)$$

$$Cele_{min} \leq \sum_{\forall_i} c_i x_i \leq Cele_{max} \quad \forall\, x_i \in \{ELE\}\ (3)$$

$$Ccp_{min} \leq \sum_{\forall_i} c_i x_i \leq Ccp_{max} \quad \forall\, x_i \in \{CP\}\ (4)$$

$$x_i \in \{0,1\} \quad (5)$$

A pesar de que la relajación lineal representa adecuadamente el problema de selección de asignaturas, es posible que subestime las asignaturas que tienen un valor esperado menor al del promedio ponderado acumulado del estudiante.

9.7 Métodos de solución

Debido a la complejidad del problema original y según la revisión bibliográfica realizada se propone usar como método de solución la meta-heurística: Algoritmos genéticos. Las ventajas de usar este método de solución son principalmente:

- **Facilidad de programación:** Es un método sencillo de formular y programar.
- **Flexibilidad:** La descripción del algoritmo genético permite adaptarlo fácilmente al proceso de selección de asignaturas.
- **Desarrollos en R:** Como se mencionó anteriormente existe el paquete Rgenoud, el cual ya tiene programado la secuencia y funciones del algoritmo genético.
- **Variedad de respuestas:** Debido a la naturaleza evolutiva del algoritmo, es posible presentar varias propuestas de conjuntos de asignaturas con valores de función objetivo similares.

9.7.1 Algoritmo genético para la solución del proceso de selección de asignaturas

Cromosoma: será un vector con un número de componentes igual al número de asignaturas candidatas según el registro histórico del estudiante. Cada uno de los componentes del vector será representado por un 1 ó 0 dependiendo si la asignatura en dicha posición estará o no en el conjunto de asignaturas sugeridas.

Función objetivo: Promedio académico ponderado acumulado esperado.

Penalización: Aquellas soluciones infactibles tomarán un valor de función objetivo arbitrariamente pequeño, en este caso -1.000.

Solución inicial: Para aumentar la efectividad del modelo se incluirá como solución inicial el conjunto de asignaturas sugeridas al resolver el problema relajado descrito anteriormente.

Tamaño de la generación inicial: se generarán al azar 1000 soluciones.

Número máximo de generaciones: Como uno de los criterios de parada se tomará como 50 el número máximo de generaciones.

Generaciones de espera: Como uno de los criterios de parada se asume que si no existe mejora de la función objetivo en 10 generaciones se debe detener el algoritmo.

Operadores genéticos usados: Clonación, copiar el individuo más apto 50 veces en la siguiente generación. Mutación uniforme, seleccionar 50 individuos aleatoriamente y reconfigurarlos usando valores al azar. Cruce simple, seleccionar 50 veces dos individuos al azar y cruzarlos de manera aleatoria para crear otro individuo.

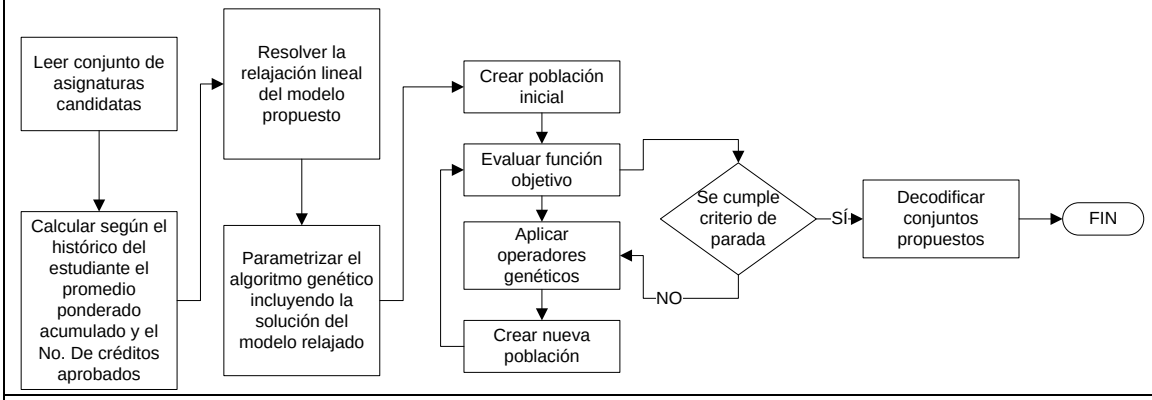
9.8 Continuación funciones creadas en el entorno de programación R statistics

Se presenta a continuación la función creada para soportar el proceso de selección de asignaturas.

9.8.1 Apartado solución algoritmo genético

Apartado dirigido a la procesamiento del algoritmo genético destinado a resolver la modelación del proceso de selección de asignaturas.

9.8.1.1 Función: Estimación de conjuntos de asignaturas sugeridos

Apartado: Aplicación algoritmo genético.
Función: Estimación de conjuntos de asignaturas sugeridos
Hipervínculo: Código - Invocación
Sintaxis: EstimarSubconjunto(DatosEstudianteRequisitosCumplidos, conjuntospropuestos, Mincred, Maxcred, MinELE, MaxELE, MinCP, MaxCP)
Parámetros: Registro histórico del estudiante, número de propuestas sugeridas, mínimo número de créditos totales a inscribir, máximo número de créditos totales a inscribir, mínimo número de créditos de asignaturas complementarias a inscribir, máximo número de créditos de asignaturas complementarias a inscribir, mínimo número de créditos de asignaturas electivas a inscribir, máximo número de créditos de asignaturas electivas a inscribir.
Función requerida: Cálculo del valor esperado de la calificación de asignaturas candidatas.
Salida: Tabla que indica los conjuntos de asignaturas sugeridos para que el estudiante los considere en el próximo periodo académico. De igual manera se indica para cada conjunto la calificación esperada de cada asignatura y el promedio ponderado acumulado esperado si se elige dicho conjunto.
Objetivo: Ofrecer al estudiante conjuntos de asignaturas que debe considerar al momento de seleccionar las asignaturas que va a inscribir en el siguiente periodo académico. Con la automatización de este proceso se pretende que el estudiante pueda disponer del tiempo necesario para analizar detalladamente los conjuntos de asignaturas que ha seleccionado personalmente y los presentados por esta metodología para crear criterios que logren obtener la mejor alternativa posible según su situación académica y convicciones personales.
Precondiciones: La función Cálculo del valor esperado de la calificación de asignaturas candidatas brinda los valores esperados acertados para cada una de las asignaturas candidatas. Los criterios de convergencia del algoritmo genético son suficientes para obtener una buena solución.
Diagrama de Flujo: 
Visualización salida:

	Asignaturas	Nota Esperada	Naturaleza	Créditos	IDsugerido
	Pensamiento Algorítmico	3.8	NFFyENF	3	4206
	Topografía y Fotogrametría	3.8	NFFyENF	2	4180
	Materiales de Construcción	4	NFFyENF	3	4033
	Laboratorio de Materiales	3.7	NFFyENF	2	4030
	Complementaria DPT-LENGS	3.9	CP	2	<NA>
	Electivas	5	ELE	3	<NA>
Promedio	Ponderado Esperado	4.12	-	-	-
Promedio	Semestre Esperado	4.08	-	-	-
	Total Créditos	15	-	-	-

Asignaturas	Nota Esperada	Naturaleza	Créditos	IDsugerido
Física Mecánica	3.4	NFFyENF	3	1340
Topografía y Fotogrametría	3.8	NFFyENF	2	4180
Materiales de Construcción	4	NFFyENF	3	4033
Laboratorio de Materiales	3.7	NFFyENF	2	4030
Complementaria DPT-LENGS	3.9	CP	2	<NA>
Electivas	5	ELE	3	<NA>
Promedio Ponderado Esperado	4.08	-	-	-
Promedio Semestre Esperado	4	-	-	-
Total Créditos	15	-	-	-

Asignaturas	Nota Esperada	Naturaleza	Créditos	IDsugerido
Cálculo Integral	3.4	NFFyENF	3	1297
Pensamiento Algorítmico	3.8	NFFyENF	3	4206
Topografía y Fotogrametría	3.8	NFFyENF	2	4180
Materiales de Construcción	4	NFFyENF	3	4033
Laboratorio de Materiales	3.7	NFFyENF	2	4030
Complementaria DPT-LENGS	3.9	CP	2	<NA>
Electivas	5	ELE	3	<NA>
Promedio Ponderado Esperado	4.06	-	-	-
Promedio Semestre Esperado	3.97	-	-	-
Total Créditos	18	-	-	-

Ejemplo:
// EstimarSubconjunto(DatosEstudianteRequisitosCumplidos, 3, 15, 20, 2, 5, 2, 4)

10 Análisis de resultados

Esta sección describe el análisis de los resultados obtenidos por la metodología para el ejemplo mencionado en la sección 9.3.1.

10.1 Caracterización de la asignatura Ecuaciones Diferenciales

De los resultados de la función 9.5.1.1 'Información estadística básica de la asignatura' se puede observar que según los valores de sus percentiles, el 50% de los datos están concentrados alrededor de 3.0 y 3.6. Su coeficiente de asimetría indica que existe una mayor frecuencia de observaciones en valores mayores a la media, sin embargo aproximadamente uno de cada cuatro estudiantes pierden la asignatura. Por otra parte su coeficiente de kurtosis sugiere que hay una cantidad considerable de datos extremos. Una apreciación que se puede obtener por los valores de la desviación estándar y el promedio es que es poco probable obtener calificaciones mayores a 4. Todas estas conclusiones se pueden apreciar de manera gráfica en la función 9.5.2.2 'Gráfico de histograma y FDP', en ella podemos detallar la distribución exacta y estimada del comportamiento de los datos. La tabla resultante de la función 9.5.2.1 'Matriz de probabilidad por asignatura' presenta el coeficiente de *Kolmogorov-smirnov* para esta asignatura y la distribución Beta con parámetros 6.69 y 4; el valor de la significancia del estadístico es 0, lo cual implica que no es apropiado estimar el conjunto de datos con esta distribución, y claramente se puede observar que se cumple lo mencionado en las precondiciones de esta función. Existe una gran concentración de registros en 3.0 por lo que no es posible estimar esta distribución a través de una función Beta, ya que se subestima las probabilidades calculadas. Según las comparaciones del porcentaje de datos bajo 3 de la función 9.5.1.1 y el resultado de la función 9.5.2.3 'Cálculo de las probabilidades estimadas' es

posible obtener el error aproximado de la medición realizada con la función de probabilidad (40%) y la proporción de casos favorables (23%).

10.2 Tendencias de la asignatura Ecuaciones Diferenciales.

La tabla resultante de la función 9.5.1.2 'Promedio por periodo y asignatura' al igual que la gráfica de la función 9.5.3.2 'Gráfico de control por asignatura y números de clase' revelan que existe una tendencia al acenso en las calificaciones obtenidas en la asignatura Ecuaciones diferenciales en los últimos 4 periodos académicos. También existe una disminución en el número de clases que están por fuera de los límites de control.

Pero, por otra parte, el gráfico que presenta la función 9.5.3.1 'Gráfico de cajas de control por asignatura y periodo' demuestra que no existe un criterio estándar en el proceso de calificación de la asignatura en el último periodo, ya que los resultados de cada número de clase son marcadamente diferentes.

También al analizar la tabla provista por la función 9.5.3.3 'Ranking de asignaturas con números de clase fuera de control' se evidencia que existen dos números de clase por fuera de los límites en el periodo 1210 en Ecuaciones Diferenciales. Estos números de clase se observan en el gráfico de la función 9.5.3.1: 2657 y 2663.

10.3 Agrupación de asignaturas

En la tabla proveniente de la función 9.5.4.2 'Agrupación por K medias' y el gráfico de la función 9.5.4.3 'Gráfico agrupación por K medias' se observa que la asignatura Ecuaciones diferenciales se encuentra en el grupo 2 con las variables: Porcentaje de calificaciones bajo 3.0 y desviación estándar, y grupo 3 con las variables: Número de créditos y mediana de la calificación. Esta primera agrupación indica que esta asignatura recae sobre aquellas que tiene una desviación estándar media-alta (0.5-0.8) y que a su vez tienen un alto porcentaje de estudiantes que la reprueban. Este tipo de caracterización señala que la asignatura tiene un gran porcentaje de estudiantes que la reprueba, pero, de igual manera, existen varios estudiantes que obtienen notas medianamente buenas (3.5-4). Mientras tanto, la segunda agrupación sólo establece que la asignatura se caracterizó por su número de créditos y su valor de la mediana concentrada cercana a 3.

10.4 Análisis de asignaturas re-cursadas

La función 9.5.5.1 'Información de asignaturas re-cursadas' y la función 9.5.1.2 'Gráfico FDP y tendencia de asignaturas re-cursadas' manifiestan que la asignatura Ecuaciones Diferenciales ha sido cursada un número máximo de 3 veces con calificaciones que aumentan según el número de veces que se repita. La gráfica de tendencia de la función del apartado 9.5.1.2 expone que una gran cantidad de estudiantes que reprueban la asignatura por primera vez tienden a perderla una segunda vez. Por otra parte la tabla generada por la función 9.5.5.3 'Posibles estudiantes retirados del programa académico' estipula que es posible que 5 estudiantes se hayan retirado del programa académico debido a conflictos con esta asignatura.

10.5 Análisis de relaciones entre asignaturas

Según la tabla generada por la función 9.5.6.2 ‘Matriz de correlaciones y asignaturas con mayor relación lineal’ las principales asignaturas que tienen relación con Ecuaciones Diferenciales son: Fluidos y Termodinámica, Pensamiento Algorítmico, Estática, Mecánica de Fluidos, Cálculo integral, Investigación de Operaciones y Física Mecánica. Dependiendo de cuál asignatura esté primera según el plan de estudios, esta relación indica una dependencia hacia o desde otras asignaturas. Por otro lado, el gráfico generado por la función 9.5.6.3 ‘Gráfico de correlación de asignaturas’ revela que no hay una relación fuerte entre las asignaturas Ecuaciones Diferenciales, Álgebra Lineal y Procesos Constructivos.

10.6 Identificación de pre-requisitos

En la tabla resultante de aplicar la función 9.5.7.1 ‘Requisitos por asignatura’ se observa que es necesario aprobar las asignaturas Cálculo Diferencial, Álgebra Lineal y Cálculo integral para poder cursar Ecuaciones Diferenciales. Esta tabla también menciona que se necesita al menos dos periodos académicos para poder cursar estas asignaturas.

10.7 Identificación de asignaturas candidatas y calificación esperada

Dado el conjunto de asignaturas aprobadas hasta el momento, el estudiante está en capacidad de cursar todas las asignaturas provistas en la tabla generada por la función 9.5.7.2 ‘Selección de asignaturas candidatas’ y 9.5.8.1 ‘Cálculo del valor esperado de la calificación de asignaturas candidatas’. Esta última presenta también los valores esperados de las calificaciones y la naturaleza de cada asignatura. Como se mencionó en los apartados anteriores esta metodología agrupa las electivas y complementarias cursadas por el estudiante.

10.8 Generación de sugerencias

La tabla concebida por la función 9.8.1.1 ‘Estimación de conjuntos de asignaturas sugeridos’ muestra las propuestas generadas por la metodología, en ella se presentan los nombres de las asignaturas, su naturaleza y código, el cual es necesario al momento de inscribir formalmente las asignaturas. También se incluye el promedio ponderado esperado del semestre y el promedio ponderado acumulado esperado si se inscriben esas asignaturas. De igual manera para evitar conflictos de intereses y sesgar las convicciones de aprendizaje del estudiante, se menciona únicamente el término ‘electivas’ y el departamento que brinda las asignaturas complementarias elegidas por el estudiante con su respectiva calificación estimada. Estas propuestas deben ser analizadas junto con las demás opciones que el estudiante seleccione y obtener un consenso unificando sus criterios y convicciones, y el acompañamiento y experiencia del consejero académico.

11 Uso de las funciones creadas en *R statistcs* e ingreso de información.

Para la aplicación de esta metodología es necesario contar con los dos paquetes de software indicados en la sección 9.2: *Microsoft Excel y R statistics.* Para la descarga e instalación de R se recomienda revisar el link presentado en la sección 9.2.2.2.

A continuación se describe cómo debe ser el procedimiento para el ingreso de información y la manipulación de las rutinas creadas en R.

11.1 Ingreso de información

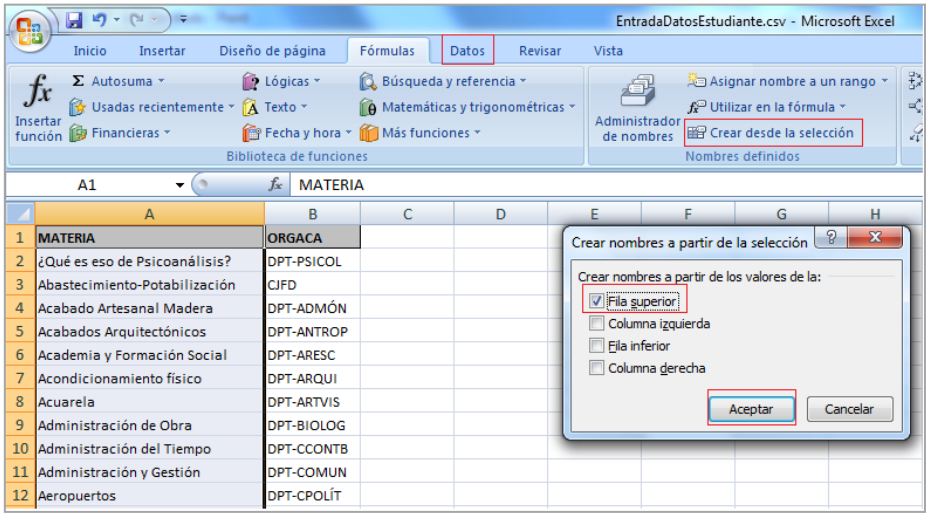
En los archivos anexos se encuentra un ejemplo de cada uno de las entradas de información que tiene esta metodología. Estas plantillas deben utilizarse como único método para el ingreso de datos a R. Como se puede observar cada una de las casillas permite únicamente ingresar los datos correspondientes a las asignaturas y organizaciones académicas que reposan en la base de datos.

Para permitir que el código generado en R cargue las entradas de información se tiene que crear una carpeta en el escritorio con el nombre 'Asignaturascsv' y dentro de ella introducir la base de datos, matriz de pre-requisitos, requisitos especiales y registro histórico del estudiante obligatoriamente con los nombres dispuestos para estos archivos en los anexos digitales. Cada uno de los archivos debe abrirse en Excel y guardarse como CSV.

11.1.1 Creación de plantilla de ingreso de registro de datos del estudiante

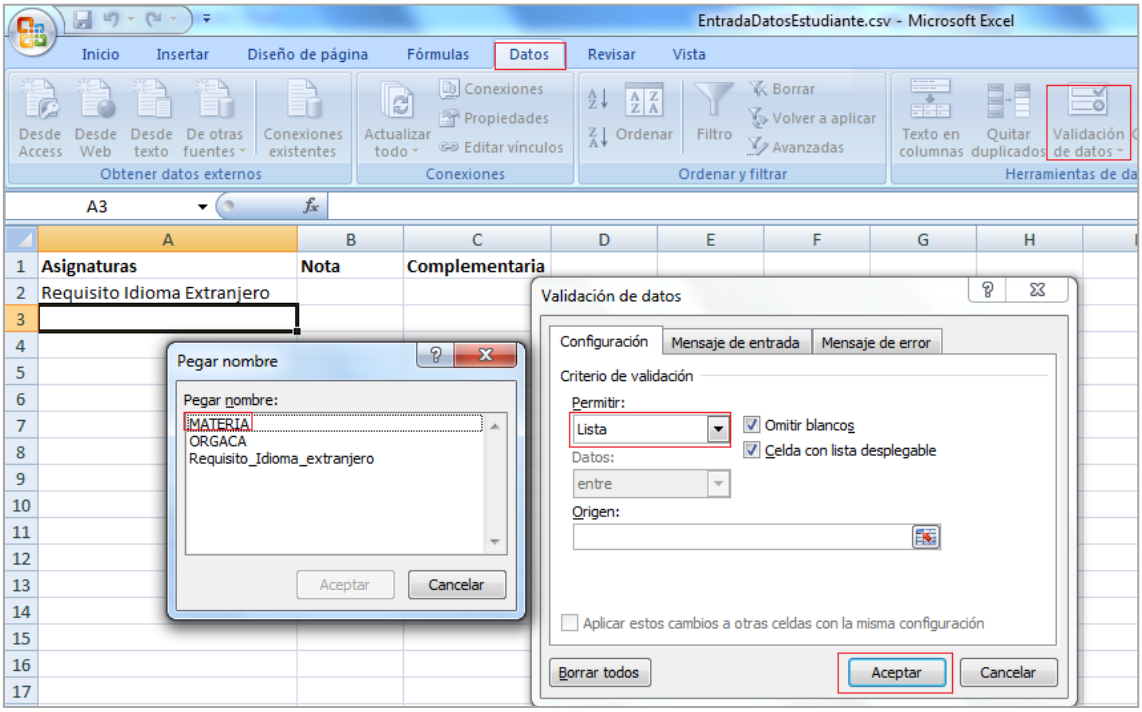
Para la replicación de la plantilla del archivo Entrada Datos Estudiante es necesario seguir los siguientes pasos:

1. Copiar todas las asignaturas y organizaciones académicas de la base de datos académica depurada en una hoja de Excel diferente de la primera.
2. Con la función Eliminar Duplicados disponible en la pestaña Datos obtener la lista única de asignaturas y organizaciones académicas.
3. Con la función Ordenar y Filtrar de la pestaña de inicio ordenar las listas alfabéticamente.
4. Seleccionar la lista de asignaturas, desde su título hasta la última asignatura y con ayuda de la función Crear desde Selección disponible en la pestaña Fórmulas darle un nombre a la lista.



5. En la primera hoja del libro se deben escribir los títulos de la tabla de ingreso de información tal y como se menciona en la sección 9.1.1.3. De igual manera la celda A2 debe tener el nombre 'Requisito de Idioma Extranjero'.

6. Para restringir los posibles nombres que se pueden introducir dentro de las casillas de la columna Asignatura se debe utilizar la función Validación de Datos disponible en la pestaña Datos. De las opciones disponibles en Permitir se debe seleccionar lista, ubicar el cursor sobre el campo Origen, oprimir la tecla F3 y seleccionar el nombre de la lista anteriormente creada que corresponda. En la figura presentada a continuación se ha agregado una lista llamada Requisito_Idioma_extranjero que contiene los valores Sí y No para evadir errores ortográficos y evitar problemas por usos de letras mayúsculas.



7. Copiar y pegar la validación en las demás casillas de la columna Asignaturas utilizando la función Pegar Validación disponible en las opciones de pegado especial.

8. Guardar el archivo con el mismo nombre como CSV en la carpeta 'Asignaturascsv'.

11.1.2 Creación de plantilla de ingreso de requisitos especiales

Utilizando el mismo concepto del punto anterior, es necesario estandarizar los nombres de las asignaturas que se ingresan a la plantilla de Requisitos Especiales de Asignaturas.

11.1.3 Manejo de asignaturas cooterminales.

Las asignaturas cooterminales son aquellas que tienen que cursarse exactamente el mismo periodo académico. Para cumplir con estas condiciones, se recomienda dos alternativas:

1. Seleccionar del total de conjuntos sugeridos sólo aquellos en los que no se violen las condiciones exigidas por las asignaturas cooterminales.

2. Fusionar las asignaturas cooterminales en una sola antes de ingresarla en el proceso automatizado de selección de asignaturas del apartado 9.5.2, para ello, ésta asignatura

debe satisfacer las siguientes condiciones: el total de créditos será la suma de los créditos de cada asignatura que lo conforma y la calificación esperada será el promedio ponderado de las calificaciones presentadas por la función del apartado 9.5.8.1 y el número de créditos respectivo.

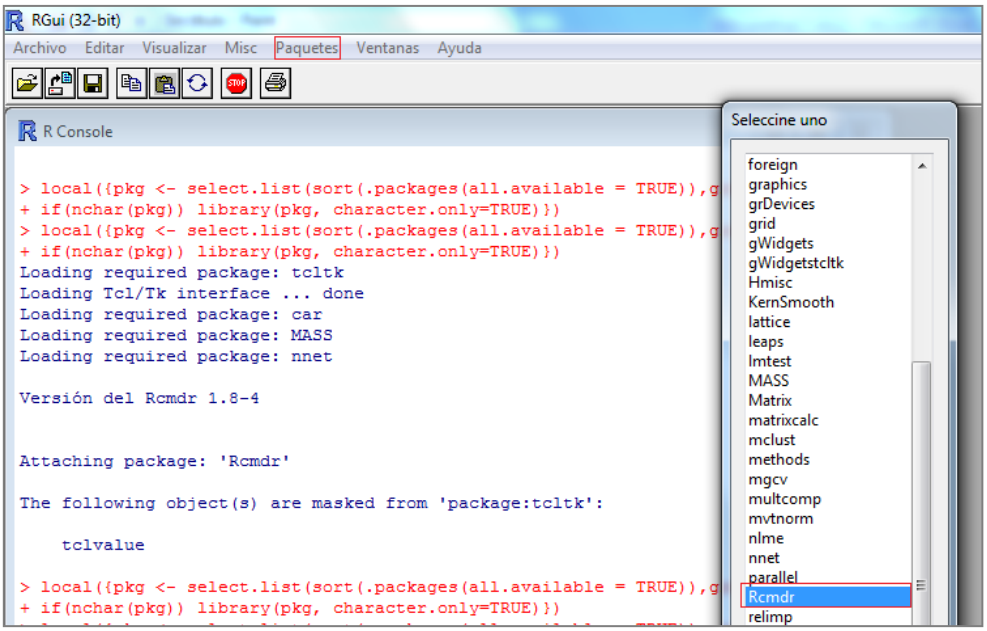
11.1.4 Instalación de paquetes en R

Después de haber instalado satisfactoriamente R es preciso añadirle los paquetes mencionados en la sección 9.2.2.3. Para ello se debe

- Ir la pestaña Paquetes y dar clic en Instalar Paquetes
- Seleccionar el país de origen de la lista desplegada
- Buscar e instalar paquetes

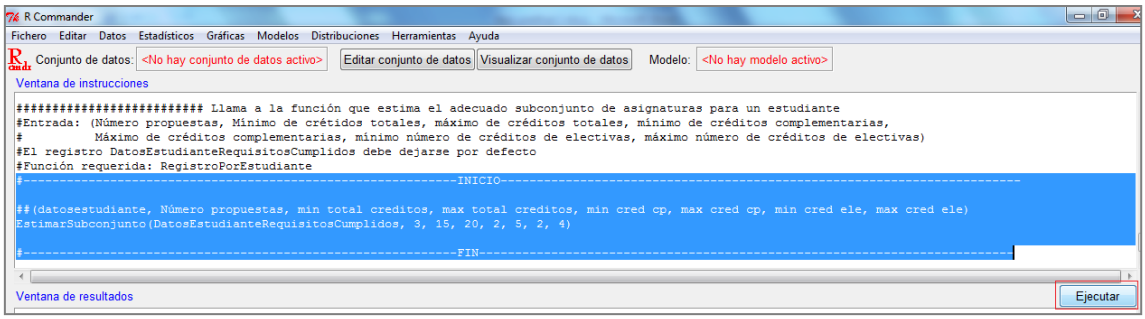
11.1.5 Compilación del código en R

Al abrir el programa R statistics debemos dirigirnos hasta la pestaña cargar paquetes y seleccionar Rcmdr, el cual abrirá una interfaz gráfica más amable para introducir el código.



Después se debe pegar el código que está anexo en un archivo de texto bajo el nombre Código R, a continuación seleccionar desde el inicio de éste hasta que está escrito el mensaje “FIN Carga de Funciones” y dar clic en el botón ejecutar. Esta acción es la responsable de cargar todas las bases de datos y las funciones creadas en R por lo puede tardar unos segundos.

Posteriormente, se debe dirigir al apartado ‘Uso de funciones’ y según la función que se requiera ejecutar se tiene que primero ingresar los parámetros pertinentes y seleccionar de inicio a fin para después con el botón ejecutar obtener el resultado.



11.1.6 Detalles del archivo 'código en R'

El archivo está ordenado de la siguiente manera: Carga de información, lectura de funciones y aplicación de las funciones creadas. Cada una de estas tres partes está separada por un aviso que indica el inicio y final de ese apartado. De igual manera también se muestra el inicio y final del código de creación de la función y de su código de invocación, el cual, presenta para cada una de ellas cuáles son los requerimientos de entrada y qué funciones se deben ejecutar antes. Si el usuario de la metodología requiere conocer al detalle el proceso de funcionamiento de alguna función en especial puede guiarse por los comentarios que indican paso a paso su estructura.

12 Trabajos futuros

El propósito fundamental del desarrollo de esta metodología consiste en el aprovechamiento de los recursos universitarios, ente ellos las bases de datos académicas para garantizar un mayor bienestar a la comunidad universitaria. Es el deber de las personas que hemos estado relacionadas con la pedagogía orientar a las nuevas generaciones hacia el camino de la excelencia usando todas las herramientas que tengamos a disposición.

El alcance de la metodología propuesta en el presente trabajo de grado está limitada al tiempo y recursos disponibles de los autores del proyecto, pero durante su desarrollo se han generado grandes ideas que podrían ser de mucha utilidad para todos los actores que intervienen en el proceso educativo, desde la facultad en general, directores de carrera, profesionales de retención, profesores, estudiantes y padres de familia.

A manera de síntesis se presentan a continuación algunas de las ideas que han surgido del análisis del desarrollo de la metodología y de la experiencia propia de necesidades requeridas y docencia:

Crear perfiles de asignaturas: Identificar cuáles son las variables más aptas para calificar y agrupar las asignaturas para así poderlas caracterizar según su: Perfil de riesgo, dificultad, aporte al programa académico y cualquier otro tipo de objetivo que requiera de un conglomerado de asignaturas.

Modelación de retiro de estudiantes: Modelar a través de procesos Markovianos, indicando las cadenas de sucesos y las probabilidades asociadas a reprobar una asignatura un número determinado de veces y su implicación en la continuidad del estudiante en el programa académico.

Permanencia en el programa académico: Recolectar información intrínseca a cada estudiante como variables relacionadas con el estilo su estilo de vida, edad, región, hobbies, ingresos, y

demás variables que podrían ayudar a modelar a través de una regresión logística, la probabilidad que tiene un estudiante de retirarse del programa académico antes de terminarlo. Este análisis podría generar alertas tempranas para que la facultad refuerce los programas de retención de estudiantes.

Refinamiento de la metodología propuesta: Incluir en las bases de datos académicas información relacionada con el estudiante antes de ingresar a la institución educativa. Recolectar variables como notas en el colegio, edad, categoría de colegio, resultado de pruebas del ICFES, podrían refinar la metodología propuesta en el presente trabajo de grado.

Mejoramiento del modelo de regresión: Al modelo de regresión propuesto se le debe incluir información relacionada con el número total de créditos inscritos en cada periodo académico y de igual manera medir el impacto que puede tener el haber repetido una asignatura un número determinado de veces en el valor esperado de la calificación de una asignatura en particular.

Mejoramiento de los procesos de control: Añadir variables que describan cada uno de los números de clase de una asignatura, estas variables pueden contener información relacionada con el número total de estudiantes inscritos, horario y docente.

Modelos de recomendación: Crear a través de técnicas estadísticas, recomendaciones de asignaturas electivas y complementarias según el perfil del estudiante y los registros históricos de las bases de datos.

Malla curriculares estándar: Crear un modelo matemático que sugiera una malla curricular particular según los ideales del estudiantes. Estos ideales pueden estar relacionados con: Pagar el mínimo dinero en mátriculas académicas, terminar el programa académico lo antes posible, homogeneizar la dificultad de cada semestre, etc.

13 Conclusiones

1. El apartado descriptivo es fundamental para poder caracterizar el comportamiento de las calificaciones de una asignatura.

2. El proceso de análisis de la base de datos académica es crítico dentro del proceso de selección de asignaturas, ya que los modelos de optimización creados funcionarán adecuadamente sólo si tienen parámetros de entrada pertinentes.

3. Es posible realizar los procesos de control de las asignaturas y números de clase usando la información provista por la base de datos académica, identificando comportamientos fuera de control en el promedio de las calificaciones de los diferentes números de clase.

4. Se sugiere modificar las consultas a través de las cuales se generan las bases de datos, debido a que existen oportunidades de añadir más información a la metodología, en particular son de interés datos asociados con docentes, horarios e información intrínseca al estudiante como, resultados de evaluación del examen de estado y edad.

5. Con la información resultante del análisis de asignaturas re-cursadas es posible obtener un estimado del número de estudiantes que se retiran del programa.

6. El apartado de análisis de requisitos puede convertirse en una herramienta útil para disminuir el tiempo usado en seleccionar las asignaturas candidatas y en verificar si un estudiante está en condiciones académicas de cursar una asignatura.

7. Para cualquier análisis relacionado con la inscripción de asignaturas es estrictamente necesario agrupar las asignaturas electivas y complementarias para así evitar que el estudiante sacrifique sus convicciones de aprendizaje por priorizar objetivos diferentes.

8. El valor esperado de la calificación del conjunto de asignaturas candidatas puede servir como alerta para que los estudiantes identifiquen antes de que empiece el periodo académico, aquellas asignaturas en las cuales existe una mayor probabilidad de obtener una baja calificación.

9. Los conjuntos de asignaturas estimados a través de esta metodología deben estar analizados en conjunto con las propias elecciones de los estudiantes y el seguimiento y apoyo del consejero académico para así garantizar que se seleccione el conjunto más adecuado según la situación académica del estudiante y su convicción de aprendizaje.

14 Referencias

[1] J. Restrepo, «El sistema de créditos académicos en la perspectiva Colombiana y MERCOSUR: Aproximaciones al modelo europeo,» *Revista de la educación superior,* pp. Vol.34 pp131-152, 2005.

[2] Vicerrectoría Académica, Implementación de un sistema de créditos académicos, una estrategia de organización de los planes de estudio, Bogotá: Pontificia Universidad Javeriana, 2002.

[3] H. Guruler, A. Istanbullu y M. Karahasan, «A new student performance analysing system using knowledge discovery in higher educational databases,» *Comput Educ,* vol. 55, nº 8, pp. 247-254, 2010.

[4] S. Martello y P. Toth, «Knapsack Problems: Algorithms and compute implementations,» *ISBN: 0-471-9240-2,* 1990.

[5] H. Keller, U. Pferschy y D. Pisinger, «Knapsack problems,» *Springer-Verlag Berlín Heidelberg,* vol. 1, nº 40286, pp. 540-570, 2004.

[6] Vicerrectoría Académica, Consejería Académica, propuesta para el consejo académico, Bogotá (DC: Pontificia Universidad Javeriana, 2002).

[7] St michaels university school, «Course selection guide,» 2013.

[8] College of arts and science, «Freshman course selection,» http://www.college.upenn.edu/freshman-courses, Fecha consulta: 14-4-2013.

[9] D. Pisinger, «Where are the hard knapsack problems?,» *Comput Oper Res,* vol. 9, nº 32(9), pp. 2271-2284, 2005.

[10] M. Babaioff, N. Immorlica, D. Kempe y R. Kleinberg, «A knapsack Secretary problem with applications,» *Microsoft research,* vol. 9, pp. 033-165, 2005.

[11] C. Bazgan, H. Hugot y D. Vanderpooten, «Implementing an efficient fptas for the 0–1 multi-objective knapsack problem,» *Eur J Oper Res,* vol. 1, nº 198, pp. 47-56, 2009.

[12] C. Bazgan, H. Hugot y D. Vanderpooten, «Solving efficiently the 0–1 multi-objective knapsack problem,» *Comput Oper Res,* vol. 1, nº 36, pp. 260-279, 2009.

[13] K. Florios, G. Mavrotas y D. Diakoulaki, «Solving multiobjective, multiconstraint knapsack problems using mathematical programming and evolutionary algorithms,» *Eur J Oper Res,* vol. 203, nº 1, pp. 14-21, 2010.

[14] K. Bretthauer y B. Shetty, «The nonlinear knapsack problem – algorithms and applications,» *Eur J Oper Res,* vol. 5, nº 138, pp. 459-472, 2002.

[15] C. D'Ambrosio y S. Martello, «Heuristic algorithms for the general nonlinear separable knapsack problem,» *eur op res,* vol. 38, nº 2, pp. 505-513, 2011.

[16] S. Kosuch y A. Lisser, «On two-stage stochastic knapsack problems,» *Discrete Applied Mathematics,* p. In Press, 2010.

[17] C. Papadimitriou, Combinatorial optimization: Algorithms and complexity, vol. 468, Prentice Hall, 1998.

[18] N. Galimyanoba, «Experimental investigations of combined algorithms of branch and bound method and dynamic programming method for knapsack,» *Journal of Computer and Systems Sciences International,* vol. 47, nº 3, p. 422–428, 2008.

[19] H. Graham y D. Joux, «New generic algorithms for hard knapsacks,» *eurocrypt,* p. 235–256, 2010.

[20] M. Dyer, W. RIha y J. Walker, «A hybrid dynamic programming/branch-and-bound algorithm for the multiple-choice knapsack problem,» *J Comput Appl Math,* vol. 58, nº 3, pp. 43-54, 1995.

[21] P. Toth, «Dynamic programming algorithms for the zero–one knapsack problem,» *Computing,* vol. 25, p. 29–45, 1990.

[22] S. Balev, N. Yanev, A. Fréville y R. Andonov, «A dynamic programming based reduction procedure for the multidimensional 0–1 knapsack problem,» *Eur J Oper Re,* vol. 186, nº 4, pp. 63-76, 2008.

[23] S. Martello, D. Pisinger y P. Toth, «New trends in exact algorithms for the 0–1 knapsack problem,» *Eur J Oper Res,* vol. 123, nº 6, pp. 325-332, 2000.

[24] J. Gorski, L. Paquete y F. Pedrosa, «Greedy algorithms for a class of knapsack problems with binary weights,» *Computers and operations research.,* 2011.

[25] F. Hillier, Handbook of metaheuristics, Kluver Academic Publishiers. , 2003.

[26] M. Alves y M. Almeida, «MOTGA: A multiobjective Tchebycheff based genetic algorithm for the multidimensional knapsack problem,» *Comput Oper Res,* vol. 34, nº 11, pp. 3458-3470, 2007.

[27] R. Beausoleil y G. Baldoquin, «Multi-start and path relinking methods to deal with multiobjetive knapsack problems,» *Ann Oper Res,* vol. 157, p. 105–133, 2008.

[28] R. Chen, Y. Huang y M. Hsien, «Solving unbounded knapsack problem based on quantum genetic algorithms,» *ACIIDS,* p. 339–349, 2010.

[29] S. Leung, D. Zhang, C. Zhou y T. Wu, «A hybrid simulated annealing metaheuristic algorithm for the two-dimensional knapsack packing problem,» *Comput Oper Res,* In Press.

[30] F. Lin, «Solving the knapsack problem with imprecise weight coefficients using genetic algorithms,» vol. 185, nº 2, pp. 133-145, 2008.

[31] D. Montgomery y G. Runner, Engineering statistics, quinta edición, Limusa, 2010.

[32] D. Montgomery, Design and analysis of experiments, segunda edición, Limusa, 2006.

[33] Brown, «A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet,» *Comput Methods Programs Biomed,* vol. 65, nº 6, pp. 191-200, 2001.

[34] N. Brauner, «Regression diagnostic using an orthogonalized variables based stepwise regression procedure,» *Computers and chemical engineering,* pp. 327-330, 1999.

[35] E. Nuñez, E. Steyerberg y J. Nuñez, «Regression Modeling Strategies,» *Rev Esp Cardio,* vol. 64, nº 6, pp. 501-507, 2011.

[36] M. Bazaraa, linear programming and network flows, segunda edición, Limusa, 2006.

[37] H. Marchand, A. Martin, R. Weismantel y L. Wolsey, «Cutting planes in integer and mixed integer programming,» *Discrete Applied Mathematics,* vol. 123, nº 11, pp. 397-446, 2002.

[38] M. Grötschel y G. Nemhauser, «George Dantzig's contributions to integer programming,» *Discrete Optimization,* vol. 5, nº 2, pp. 168-173, 2008.

[39] C. Wilbaut y S. Hanafi, «New convergent heuristics for 0–1 mixed integer programming,» *Eur J Oper Res,* vol. 195, nº 5, pp. 62-74, 2009.

[40] C. Yeh, C. Chu y K. Wu, «Molecular solutions to the binary integer programming problem based on DNA computation,» *BioSystems,* vol. 83, nº 1, pp. 56-66, 2006.

[41] M. Sallán y A. Suñé, «Métodos cuantitativos de organización industrial,» *I.UPC,* pp. 84-89, 2002.

[42] J. Teghem, T. El-Ghazali y J. Wiley, «Metaheuristics From Design to Implementation,» *Eur J*

Oper Res, vol. 250, nº 9, pp. 486-487, 2010.

[43] F. Glover y G. Kochenberger, Handbook of metaheuristics, Kluwer Academic Publisher, 2003.

[44] D. Fayyad, «Knowledge Discovery in Database: An overview,» *Microsoft Research.*

[45] C. Morita y H. Tsukimoto, «Knowledge discovery from numerical data,» *Knowledge-Based Syst,* vol. 10, nº 5, pp. 413-419, 1998.

[46] U. Fayyad y P. Stolorz, «Data mining and KDD: Promise and challenges,» *Future Generation Comput Syst,* vol. 11, nº 2, pp. 99-115, 1997.

[47] J. Velosa, Guía de Excel Básico, Universidad EAN, 2011.

[48] TeamR Development, «Introducción a R,» Versión 1.0.1, 2000.

[49] W. Owen, «The R guide,» University of Richmond, Versión 2.5, 2010.

[50] J. Fox, L. Andronic, M. Ash, M. Bouchet y T. Boye, «Package Rcmdr,» 2013.

[51] W. Mebane y J. Singh, «R version of GENetic Optimization Using Derivatives,» 2013.

[52] K. Homik y S. Theussl, «R/GNU Linear Programming Kit Interface,» 2012.

[53] J. Keats, «A theorical distribution for mental test scores,» *Psychometrika,* vol. 27, nº 1, 1962.

[54] S. Livingston y D. Lewis, «Estimating the Consistency and Accuracy of Classifications Based on Test Scores,» *Journal of Educational Measurement,* vol. 32, nº 2, pp. 179-197, 2005.

[55] P. Meyer, Introductory Probability and Statistical Applications segunda edición, Addison Wesley, 1980.

[56] L. Santibañez, «Why should care if teachers get A's: Teacher test scores and student achivement in Mexico,» *Economics of Education Review,* vol. 25, p. 510–520, 2006.

[57] N. Douglas y T. Sass, «Teacher training, teacher quality and student achievment,» *Journal of Public Economics,* 2010.