



UNIVERSIDAD CATÓLICA
de Colombia

**SOFTWARE PARA EL CÁLCULO DEL VALOR AGREGADO APORTADO A
LOS ESTUDIANTES POR LAS UNIVERSIDADES A PARTIR DE LAS
PRUEBAS SABER PRO Y SABER 11 PARA LOS AÑOS 2016 Y 2017**

**ROBERTSON SIERRA ARIAS
CAMILO ANDRÉS LEURO CAMACHO**

**UNIVERSIDAD CATÓLICA DE COLOMBIA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTÁ D.C. COLOMBIA
2019**

**SOFTWARE PARA EL CÁLCULO DEL VALOR AGREGADO APORTADO A
LOS ESTUDIANTES POR LAS UNIVERSIDADES A PARTIR DE LAS
PRUEBAS SABER PRO Y SABER 11 PARA LOS AÑOS 2016 Y 2017**

**ROBERTSON SIERRA ARIAS
CAMILO ANDRÉS LEURO CAMACHO**

Trabajo de grado presentado como requisito parcial para optar al título de:
INGENIERO DE SISTEMAS

**ASESOR:
MSc. JUAN CARLOS BARRERO CALIXTO**

**UNIVERSIDAD CATÓLICA DE COLOMBIA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTÁ D.C. COLOMBIA
2019**



La presente obra está bajo una licencia:
Atribución 2.5 Colombia (CC BY 2.5)
Para leer el texto completo de la licencia, visita:
<http://creativecommons.org/licenses/by/2.5/co/>

Usted es libre de:

- Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra
- hacer obras derivadas
- hacer un uso comercial de esta obra



Bajo las condiciones siguientes:



Atribución — Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante (pero no de una manera que sugiera que tiene su apoyo o que apoyan el uso que hace de su obra).

NOTA DE ACEPTACIÓN

Aprobado por el comité de grado en cumplimiento de los requisitos exigidos por la Facultad de Ingeniería y la Universidad Católica de Colombia para optar al título de Ingeniero de Sistemas.

Jurado

Jurado

Juan Carlos Barrero Calixto
Asesor

A Dios y a todas las personas que nos apoyaron para hacer posible este trabajo.

Contenido

LISTA DE ILUSTRACIONES	1
LISTA DE TABLAS	2
RESUMEN	3
ABSTRACT	4
1. INTRODUCCIÓN	5
1.1 Planteamiento del problema	7
1.2 Justificación	8
1.3 Objetivos	9
1.3.1 Objetivo General	9
1.3.2 Objetivos específicos	9
1.4 Delimitación	10
1.4.1 Espacio geográfico	10
1.4.2 Tiempo	10
1.4.3 Contenido	10
1.4.4 Alcance	10
1.4.5 Limitaciones	11
2. MARCO DE REFERENCIA	12
2.1 Marco Conceptual	12
2.2 Marco Teórico	15
2.2.1 Modelo Lineal Jerárquico	15
2.2.2 Los modelos lineales mixtos	16
2.2.3 Criterio de información de Akaike (Experian Commercial Risk Database AIC)	16
3. ESTADO DEL ARTE	18
4. METODOLOGÍA	19
4.1 Metodología de Desarrollo	19
5. REQUERIMIENTOS	21
5.1 Especificación de los requerimientos	21
6. DISEÑO	22
6.1 Normalización de Base de Datos	22
6.2 Diagrama de ejecución	22
6.3 Diagrama de Componentes	23
6.4 Modelo Entidad Relación	24

6.5	Diseño de los MockUps	25
7.	DESARROLLO	28
7.1	Limpieza de archivos	28
7.2	Cargue de archivos a la base de datos	29
7.2.1	Cargue de Pruebas Saber 11	29
7.2.2	Cargue de Llaves de cruce entre Saber 11 y Saber pro	32
7.2.3	Cargue de Pruebas Saber Pro	32
7.2.4	Generación de tabla Saber Pro Agregado saber 11	34
7.3	Grupo de referencia	35
7.3.1	Elección de año sobre la cual se va a realizar el cálculo.....	35
7.3.2	Elección de grupos de referencia por áreas de conocimiento del ICFES	35
7.3.3	Elección de programas según grupo de referencia por área de conocimiento del ICFES escogido	35
7.3.4	Elección de universidades según programas escogidos.....	45
7.3.5	Creación del grupo de referencia	45
7.4	Cálculo del valor agregado	45
7.4.1	Elección de la variable dependiente	45
7.4.2	Elección de las variables independientes	46
7.4.3	Cálculo del valor agregado sobre el grupo de referencia.....	46
7.5	Exportación de resultados	47
8.	PRUEBAS	49
8.1	Prueba de caja blanca:	49
8.1.1	Componentes front:	49
8.1.2	Componentes Backend	50
8.2	Prueba de caja negra:	51
9.	CONCLUSIONES.....	52
10.	RECOMENDACIONES	53
11.	TRABAJOS FUTUROS	54
12.	BIBLIOGRAFÍA	55

LISTA DE ILUSTRACIONES

Ilustración 1 - Methodology RAD.....	20
Ilustración 2 - Diagrama de ejecución del software.	22
Ilustración 3 - Diagrama de Componentes.	23
Ilustración 4 - Modelo Entidad Relación.	24
Ilustración 5 - Pantalla inicial - Calculo Valor Agregado.	25
Ilustración 6 - Pantalla Cargar Bases de Datos.....	25
Ilustración 7 - Pantalla selección año y grupo de referencia ICFES.	26
Ilustración 8 - Pantalla Selección programas según grupo de referencia ICFES. .	26
Ilustración 9 - Pantalla Selección Universidades según programas escogidos.....	27
Ilustración 10 - Pantalla Cálculo de Valor Agregado.....	27
Ilustración 11 - Resultado Modelo de regresión lineal Mixto.....	47
Ilustración 12 - Ejemplo de Modelo hecho en R Studio.	51
Ilustración 13 - Ejemplo de Modelo hecho en Python.....	51

LISTA DE TABLAS

Tabla 1 - Limpieza de caracteres.	28
Tabla 2 - Normalización de columnas.	30
Tabla 3 - Columnas finales Saber 11.	31
Tabla 4 - Nivel educativo padres – Valor.	32
Tabla 5 - Horas trabajo de estudiante – Valor.	32
Tabla 6 - Estrato Socioeconómico – Valor.	33
Tabla 7 - Género – Valor.	33
Tabla 8 - Tipo Universidad – Valor.	33
Tabla 9 - Metodología de estudio – Valor.	34
Tabla 10 - Valor promedio de matrícula.	34
Tabla 11 - Programas Académicos según Grupo de Referencia ICFES.	35
Tabla 12 - Ejemplo cálculo Valor Agregado.	48

RESUMEN

El Instituto Colombiano para el Fomento a la Educación Superior (ICFES) para medir la calidad de la educación superior en Colombia se apoya en los resultados que se obtienen a través de las pruebas Saber Pro, cuyos resultados se han realizado clasificaciones y valoraciones. En la actualidad se plantea la necesidad de medir el valor agregado que las instituciones le aportan al estudiante, por lo que no solo se considera los resultados saber Pro, sino que los modelos están considerando los resultados Saber 11 que mide las competencias que adquirió el estudiante durante su formación académica en el colegio.

Para este proyecto se especificaron los requerimientos según el estándar IEEE 830 a partir de los requisitos del modelo para el cálculo del valor agregado realizado en el trabajo de grado en Magister en Investigación Operativa y Estadística de Juan Carlos Barrero y metodología que actualmente adoptó la Universidad Católica de Colombia, para las competencias genéricas de sus ocho programas de pregrado, el cual consta de varios procesos, limpieza de bases de datos, ensamble de base de datos, formación de grupos de referencia, cálculo del valor agregado y salida tablas de clasificación, que actualmente se realizaban de forma manual, es por esta razón, en este proyecto de grado se desarrolla un Software para automatizar cada uno de los procesos realizados y minimizar el tiempo de trabajo que esto incluye. Esto nos permitirá realizar una comparación de los resultados obtenidos y trabajar de forma automática los futuros resultados teniendo en cuenta el actual formato de las pruebas desarrolladas por el ICFES.

Palabras Clave: Valor agregado, competencias genéricas, Saber Pro, automatización.

ABSTRACT

The Colombian Institute for the Encouragement of Higher Education (ICFES) to measure the quality of higher education in Colombia relies on the results obtained through the Saber Pro tests, the results of which have been classified and evaluated. At present, there is a need to measure the added value that institutions provide to students, so that not only are the Saber Pro results considered, but the models are also considering the Saber 11 results, which measure the competencies acquired by students during their academic formation at school.

For this project, the requirements according to the IEEE 830 standard were specified based on the requirements of the model for the calculation of added value carried out in the Master's degree work in Operational Research and Statistics of Juan Carlos Barrero and the methodology currently adopted by the Catholic University of Colombia, for the generic competencies of its eight undergraduate programs, which consists of several processes, database cleaning, database assembly, formation of reference groups, calculation of value added and output classification tables, currently performed manually, is for this reason, in this degree project develops a software to automate each of the processes performed and minimize the work time that this includes. This will allow us to make a comparison of the results obtained and automatically work on future results taking into account the current format of the tests developed by ICFES.

Keywords: Added value, generic competences, Saber Pro, automation.

1. INTRODUCCIÓN

El valor agregado es el aporte que da una institución de educación superior a los estudiantes que cursan sus programas académicos, este valor se puede medir a partir de las pruebas Saber que el ICFES imparte a los estudiantes que están terminando su ciclo académico.

En los resultados obtenidos por Barrero¹ en su proyecto de grado para optar por el título de Magíster en Investigación Operativa, podemos evidenciar que diversos factores y variables para el resultado de un modelo lineal aplicado al cálculo del valor agregado. Las condiciones socioeconómicas, la edad, si actualmente trabaja, entre otras variables, son determinantes para la selección de un modelo aproximado para medir el efecto aula que proporciona la institución al estudiante. Actualmente este proceso es realizado de forma manual por estadísticos e investigadores, con ayuda de herramientas como Excel, SPSS, R, entre otros, que facilitan el cálculo estadístico, pero no la limpieza de la información, el ensamble de bases de datos ni la estandarización de la información necesaria para estimar el modelo lineal jerárquico; modelo que se utiliza para hallar el valor agregado.

Dentro de las instituciones de educación se encuentran características que restan o aportan valor agregado al estudiante a lo largo de su carrera, desarrollando mejor sus habilidades, conocimientos y competencias. El estudio realizado por Bukouski y Kobus² en las escuelas públicas de Polonia, indica que la mayor parte de desinterés en el aprendizaje se debe a la falta de recursos y al sobrecupo en las aulas. Según Gallardo et al.³ se puede evidenciar que uno de los factores más influyentes en la desigualdad de las instituciones de educación superior es el valor socio-económico, entre más rentabilidad tenga

¹ BARRERO CALIXTO, Juan Carlos, Modelo de Valor Agregado para la carrera de Ingeniería de Sistemas y Computación a partir de Saber PRO 2015 y 2016. Pereira, 2018. Universidad Tecnológica de Pereira UTP, Facultad de Ciencias Empresariales

² BUKOUSKI, Pawel; KOBUS, Martina; Economics of Education Review: The threat of competition and public school performance: Evidence from Poland. December 2018. London School of Economics and Political Science, ScienceDirect URL: <https://www.sciencedirect.com/science/article/pii/S0272775717306878>

³ GALLARDO ERASO, Luis Álvaro; RINCÓN QUIÑONES, Carlos Gustavo, Pedagogía y Saberes No. 35: La desigualdad universitaria, una realidad nacida del mito de la eficiencia. Bogotá, 2011, 14 h. Universidad Pedagógica Nacional. Facultad de Educación URL: <http://revistas.pedagogica.edu.co/index.php/PYS/article/download/948/963/>

una institución, más oportunidad de desarrollo habrá para un estudiante universitario, a esto se le puede sumar la infraestructura y la calidad de aprendizaje ofrecida por cada institución de educación superior.

En el momento de calcular el valor agregado que una IES aporta a sus estudiantes, es necesario realizar una serie de procesos sobre la información de las pruebas de estado de forma manual, lo cual genera para el investigador un gasto de tiempo y esfuerzo. Esto se puede mitigar si se implementa una solución automatizada que realice los procesos rutinarios y repetitivos de limpieza y organización de la información que necesita un investigador para medir el aporte de aprendizaje que las universidades entrega a sus estudiantes.

En la Universidad Católica de Colombia, para medir el valor agregado de sus programas de pregrado, se utilizan modelos lineales jerárquicos, aplicados a un grupo específico de referencia, lo que permite comparar sus programas con los de otras instituciones. El inconveniente de este proceso es que se realiza de forma manual. Por lo que se plantea el problema: ¿Cómo se puede automatizar el proceso para el cálculo de valor agregado que las universidades aportan a sus estudiantes, aplicando el método estadístico adoptado por la Universidad Católica de Colombia, que usa los resultados de las pruebas Saber PRO cruzadas con Saber 11?

Para dar solución al problema anterior, se desarrolló un software que calcula el valor agregado que concede una institución de educación superior a los estudiantes, utilizando las bases de datos de las pruebas SABER PRO de los años 2016-2017 cruzadas con las pruebas SABER 11 suministradas por el ICFES, que después de un proceso de limpieza y normalización de datos, se ejecutará un modelo lineal jerárquico sobre ellas, cuyo resultado será una tabla de clasificación de las instituciones por carrera o facultad según el valor agregado que estos aportan a sus estudiantes, midiendo el efecto aula que las IES tienen sobre sus estudiantes.

1.1 Planteamiento del problema

El ICFES y los medios de comunicación colombianos, históricamente han usado como sistema de calificación para el escalafón de las universidades en el país los promedios simples aplicados a la prueba Saber Pro⁴. Sin embargo, existe una manera más pertinente que valora los esfuerzos al interior de cada universidad que es el escalafón aplicando el efecto aula o valor agregado que aporta cada institución a sus estudiantes, que considera el nivel de salida del colegio de los estudiantes con saber 11 y el nivel de salida con saber Pro, a través de los mismos resultados entregados por el examen de estado.

Sobre las medidas de valor agregado se han realizado varios trabajos en Colombia: está el de la universidad Nacional (2015), la Universidad Santo Tomás (2016), Universidad Tecnológica de Pereira (UTP) (2018), entre otros. La Universidad Católica de Colombia, tomo como referencia este último trabajo, que fue presentado para optar a grado de maestría y que tiene por título: “Modelo de valor agregado para la carrera de Ingeniería de sistemas y computación a partir de saber Pro 2015 y 2016.” realizado por Juan Carlos Barrero y que actualmente se está implementando para todos los programas de pregrado.

El modelo desarrollado en la UTP se hizo solo para el programa de ingeniería de sistemas 2015 y 2016, la Universidad Católica lo necesita generalizar para sus 8 programas de pregrado, y poderlo aplicar para los resultados de los años 2016, 2017 y 2018, ya realizados y para los futuros años, mientras el ICFES no cambie la estructura de la prueba.

Al considerar los procesos para aplicar el modelo, la Universidad Católica encontró que en el trabajo de la UTP, todo se había realizado de forma manual, en todos sus procesos, que son: descarga, ensamble de las bases de datos, limpieza, creación de grupos de referencia, estadística descriptiva, aplicación de los modelos lineales mixtos, cálculo del valor agregado y que terminó en creación de tablas de resultados para cada una de las 4 competencias genéricas estudiadas, proceso que informó el investigador superó las 240 horas de trabajo.

⁴ ESTEBAN, Belkys P. Educación de calidad: Se renueva la institución del Icfes. 2010. Ministerio de Educación Nacional de Colombia. URL: <https://www.mineduccion.gov.co/observatorio/1722/article-234655.html>

Estudiando la metodología aplicada en la UTP para el cálculo del valor agregado, se identifica que varios de los procesos se pueden automatizar, para simplificar futuras investigaciones y la aplicación a los 8 pregrados de la Universidad Católica.

Por lo anterior podemos realizar el siguiente planteamiento del problema: ¿Cómo se puede automatizar el proceso para el cálculo de valor agregado que las universidades aportan a sus estudiantes, aplicando el método estadístico adoptado por la Universidad Católica de Colombia, que usa los resultados de las pruebas Saber PRO cruzadas con Saber 11?

1.2 Justificación

El cálculo de valor agregado con el método desarrollado en la Universidad Católica de Colombia, es un proceso dispendioso ya que inicia con la limpieza de la base de datos de Saber Pro, considerando que solo el año 2016 tiene más de 175.000 registros⁵ y 158 variables, por lo que el investigador para tener las cuatro competencias genéricas medibles con respecto al grupo de referencia puede tardar más de tres semanas, por lo que se busca a través de la automatización reducir los costos de tiempo de los investigadores de la Universidad en sus proyectos.

Automatizando el cálculo de valor agregado a través del software se busca mitigar el tiempo en cada una de las fases realizadas para el cálculo, realizando de manera automática el cargue y limpieza de las bases de datos y obteniendo los resultados que se desean en cuestión de minutos

Actualmente en Colombia no hay disponible un software que calcule directamente el valor agregado a un grupo de referencia específico, por lo que se aportaría con este aplicativo a la tecnificación del proceso.

El cálculo de valor agregado para grupos de referencia creados a la medida de las IES otorga un insumo a las directivas y al gobierno para evaluar la calidad de la educación superior del grupo de referencia, a los directores de programa

⁵ ICFES, Saber Pro: Informe Nacional de Resultados 2016-2017. Bogotá, 2018, 46h. Gobierno Nacional de Colombia. ISBN: 978-958-11-0817-6.

URL: <http://www2.icfesinteractivo.gov.co/investigacionFormulario/docman/investigadores-y-estudiantes-de-posgrado/informes-de-resultados-evaluaciones-nacionales/informes-de-saber-pro/5448-informe-nacional-de-resultados-saber-pro-2016-2017/file?force-download=1>

para ver fortalezas y debilidades que hay en la enseñanza de la profesión, y a los estudiantes un costo beneficio a la hora de escoger universidad.

1.3 Objetivos

1.3.1 Objetivo General

Automatizar el cálculo del valor agregado desarrollando un software que utilice el modelo lineal jerárquico de la Universidad Católica de Colombia aplicado a las competencias genéricas Saber Pro 2016-2017 y Saber 11, para medir el efecto aula que las IES tienen sobre sus estudiantes.

1.3.2 Objetivos específicos

- Formalizar los requerimientos funcionales a partir de la información entregada por el docente Juan Carlos Barrero e identificar requerimientos adicionales en función del software.
- Diseñar un software que automatice el cálculo del valor agregado teniendo en cuenta los requerimientos funcionales dados por la metodología estadística desarrollada en la UCC.
- Desarrollar el software que calcule el valor agregado en grupos de referencia a partir de las pruebas estandarizadas saber Pro cruzado con Saber 11.
- Realizar pruebas de funcionamiento del software y el código fuente por medio de pruebas de caja blanca y realizando comparación de los resultados del trabajo de grado de la UTP por medio de pruebas de caja negra.

1.4 Delimitación

1.4.1 Espacio geográfico

El trabajo de grado es realizado en la Universidad Católica de Colombia, con los resultados de las pruebas Saber presentadas en toda Colombia.

1.4.2 Tiempo

La duración del trabajo de grado está delimitado al tiempo que sea propuesto por el director de proyectos dentro de un periodo académico.

1.4.3 Contenido

El documento presentado contiene la información necesaria y relevante considerada por el asesor y los estudiantes para el desarrollo del software que automatice el cálculo de valor agregado, incluyendo el estado del arte.

1.4.4 Alcance

- El desarrollo del Software cumplirá con el cálculo de valor agregado siguiendo el método estadístico desarrollado en la UCC y cumplirá con las pruebas de Verificación de caja blanca y de caja negra.
- Para visualizar la información se presentará en una tabla en donde se mostrará la información de la IES del grupo de referencia que el usuario haya creado, adicionalmente la información es clasificada por valor agregado y por promedios simples
- El software permitirá la creación de grupos de referencia y realizará los cálculos de valor agregado y promedios simples sin tener en cuenta el número de IES que se hayan seleccionado.
- Para que el software posea una base de datos actualizada permitirá ingresar la información de los últimos años proporcionada por el ICFES teniendo en cuenta que la estructura de la prueba no vaya a presentar cambios
- El software presentará un ambiente intuitivo para facilitar resultados de los cálculos generados y así se pueda tomar la decisión correcta acerca de la IES más óptima de acuerdo a las preferencias seleccionadas

1.4.5 Limitaciones

- El software únicamente trabajará con la estructura y la información que poseen las bases de datos entregadas por el ICFES.
- El software utilizará las pruebas SABER PRO, cruzadas con las pruebas SABER 11 de los mismos estudiantes que presentaron las pruebas SABER PRO de los años 2016 y 2017.
- Se contemplarán y utilizarán las pruebas SABER PRO, cruzadas con las pruebas SABER 11 de los años 2016 y 2017; Las pruebas de los años anteriores no se tendrán en cuenta por que el ICFES en 2015 hizo cambios en la estructura de la prueba por lo que no son comparables con los años siguientes, además, para el momento del desarrollo de este trabajo el ICFES no habrá entregado los resultados de la prueba realizada el 7 de octubre de 2018, es por ello que no se utilizarán las bases de este año.
- El software realizará el cálculo del valor agregado utilizando las competencias genéricas de las pruebas SABER PRO, cruzadas con las pruebas SABER 11 de los años 2016 y 2017; hoy en día los investigadores no han podido calcular el valor agregado con las competencias específicas.
- El software mostrará únicamente información del grupo de referencia que se haya creado para que los usuarios realicen sus propios análisis sin intervención del aplicativo.
- Este es un software desarrollado para investigadores de valor agregado en educación superior con manejo de estadística, mas no para un público general.

2. MARCO DE REFERENCIA

2.1 Marco Conceptual

- **Aporte relativo:** El ICFES lo define como un método de refinamiento del cálculo del valor agregado, que tiene como objetivo la medición del aporte de una institución en términos de aprendizaje en comparación con otras instituciones de educación ofreciendo una serie de indicadores de calidad en la educación para cada grupo de referencia, de esta forma las diferencias entre instituciones se atribuyen a la calidad de la educación en comparación a las competencias generales que imparten.⁶
- **Bootstrap:** Es un marco HTML y CSS de plantillas y temas diseñado para ayudar a impulsar el desarrollo de aplicaciones y sitios web⁷.
- **Competencias ciudadanas:** Son un grupo de conocimientos y habilidades cognitivas, emocionales y comunicativas, que conjuntamente hacen posible que el ciudadano actúe de manera activa, solidaria y de forma constructiva en la sociedad democrática a la cual pertenece.⁸
- **Competencias específicas:** Son las pruebas diseñadas dependiendo del grupo de referencia de estudio que corresponde al pregrado que se quiere evaluar, actualmente el ICFES cuenta con 40 módulos asociados de los cuales solo se presentan las competencias orientadas al área de formación profesional.⁹
- **Competencias genéricas:** Tienen como objetivo evaluar y proporcionar un reporte del grado de desarrollo de habilidades y conocimientos generales de estudiantes que aprobaron el 75% de los créditos de sus respectivos programas de formación universitaria profesional. Estas Competencias son habilidades intelectuales que fomentan el aprendizaje de aptitudes comunicativas, interpretativas, argumentativas y

⁶ ICFES, mejor saber – Valor agregado y Aporte relativo – Aporte Relativo. 2019. URL: <http://www.icfes.gov.co/aporte-relativo>

⁷ Bootstrap Templates & Themes from WrapBootstrap. (s.f.). Recuperado 21 abril, 2019, de <https://wrapbootstrap.com/>

⁸ Ministerio de Educación – Evaluación de Competencias Ciudadanas. 2012. Margarita Peña Borrero- ICFES. URL: https://www.mineducacion.gov.co/cvn/1665/articles-314094_archivo_pdf_6.pdf

⁹ ICFES, Guía de orientación Saber Pro – Competencias específicas. 2017. Publicación del Instituto Colombiano para la Evaluación de la Educación, pág. 11. URL: http://www.fce.unal.edu.co/media/files/Guia_de_orientacion_competencias_especificas_modulo_de_formulacion_evaluacion_y_gestion_de_proyectos_saber_pro_2017.pdf

propositivas, ya sean orientadas desde un entorno educativo a uno laboral.^{10 11}

- **Grupos de Referencia Saber Pro:** Son un conjunto de programas profesionales con características de formación semejantes que tienen combinaciones de módulos de competencias específicas similares.^{12 13}
- **Jquery:** Es una biblioteca de JavaScript rápida, pequeña y con muchas funciones. Hace que cosas como la manipulación de documentos HTML, el manejo de eventos, la animación y Ajax sean mucho más simples y fáciles de usar, funciona en una gran cantidad de navegadores¹⁴.
- **ICFES:** Es el Instituto Colombiano para el Fomento de la Educación Superior, organismo encargado de promover la educación superior en Colombia. Inicialmente se creó para evaluar a los alumnos que cursaban el último año de secundaria. Su objetivo es realizar funciones de investigación para reconocer los factores que influyen en una educación de calidad, así como brindar información para la toma de decisiones y el mejoramiento de esta.¹⁵
- **Investigación formativa:** Representa una alternativa eficaz que permite estimular las habilidades de los estudiantes mejorando la calidad y estructura de la investigación y el perfil profesional de la formación de los estudiantes.¹⁶
- **Pandas:** proporciona estructuras de datos de alto rendimiento fáciles de usar, y herramientas de análisis de datos para el lenguaje de programación Python¹⁷.
- **Pruebas Saber 11:** Es la evaluación del nivel de la Educación Media a partir del año 2014 se alinea con las evaluaciones de la Educación Básica

¹⁰ ICFES, Guía de orientación Saber Pro – Módulos de competencias genéricas. 2018. Publicación del Instituto Colombiano para la Evaluación de la Educación, pág. 9. URL: <https://bit.ly/2ZnOMt8>

¹¹ Ministerio de Educación Nacional, Educación Superior – Competencias genéricas en educación superior. 13 de diciembre del 2019. URL: https://www.mineducacion.gov.co/1621/articles-92779_archivo_pdf_Boletin13.pdf

¹² ICFES, Mejor saber – Estructura general del examen Saber Pro – Grupos de referencia Saber Pro 2016-1. URL: <http://www2.icfesinteractivo.gov.co/investigacionFormulario/terminos-de-uso/item/1934-grupos-de-referencia-saber-pro-2018-1>

¹³ Universidad ICESI – Resultados de Icesi por GRUPOS DE REFERENCIA. URL: <http://www.icesi.edu.co/saberpro2018/index.php/resultados/por-grupos-de-referencia>

¹⁴ jQuery. (s.f.). Recuperado 21 abril, 2019, de <https://jquery.com/>

¹⁵ ICFES Interactivo 2018 - ICFES, qué es y para qué sirve el examen de estado. 2012. URL: <http://icfesinteractivo.info/que-es/>

¹⁶ Procedia - Social and Behavioral Sciences Volume 176, 20 February 2015, Pages 940-945 - Formative Research in Higher Education: Some Reflections. URL: <https://www.sciencedirect.com/science/article/pii/S1877042815005996>

¹⁷ Python Data Analysis Library — pandas: Python Data Analysis Library. (s.f.). Recuperado 20 abril, 2019, de <https://pandas.pydata.org/>

y de esta forma brindar información a la comunidad educativa sobre el desarrollo de las competencias básicas que debe fomentar un estudiante en su vida escolar. Es considerada como una herramienta dedicada a retroalimentar al Sistema Educativo.¹⁸

- **Pruebas Saber Pro:** Es un instrumento estandarizado para la evaluación externa de la calidad de la educación superior. Forma parte, con otros procesos y acciones, de un conjunto de instrumentos que el Gobierno nacional dispone para evaluar la calidad del servicio público educativo y ejercer su inspección y vigilancia.¹⁹
- **Sqlalchemy:** es el kit de herramientas SQL de Python y el Mapeador relacional de objetos²⁰.
- **Statsmodels:** un módulo de Python que proporciona clases y funciones para la estimación de muchos modelos estadísticos diferentes, así como para realizar pruebas estadísticas y la exploración de datos estadísticos²¹.
- **Valor agregado:** El Instituto Colombiano para la evaluación de la educación ICFES, define valor agregado cómo: *“Es un estudio que se hace en el campo de la educación, el cual intenta medir cuánto aporta una institución a las competencias de sus estudiantes. Cuando un alumno entra a un centro educativo llega con unas habilidades previas, por lo que sus capacidades al terminar su ciclo académico no se deben solo a lo que aprendió por su paso en él, sino también a lo que sabían antes de entrar al mismo. Por este motivo surgen los estudios de VA, los cuales intentan aislar lo aprendido en una institución de las condiciones iniciales, para poder medir de una forma más precisa la calidad en la formación académica”*²²
- **Web2Py:** este es un marco de desarrollo completamente integrado, es decir, contiene todos los componentes que se necesitan para desarrollar aplicaciones web totalmente funcionales, este framework fue desarrollado en lenguaje python y permite programar las aplicaciones web en este mismo lenguaje²³

¹⁸ Ministerio de Educación Nacional - Estándares Básicos de Competencia, ICFES Saber 11°. 2018. URL: https://www.mineducacion.gov.co/1759/w3-article-244735.html?_noredirect=1

¹⁹ Pontificia Universidad Javeriana Cali, ¿Qué es Saber Pro? URL:

<https://www.javerianacali.edu.co/admisiones/saberpro/que-es-saber-pro>

²⁰ SQLAlchemy - The Database Toolkit for Python. (s.f.). Recuperado 24 abril, 2019, de <https://www.sqlalchemy.org/>

²¹ Perktold, J. O. S. E. F., Seabold, S. K. I. P. P. E. R., & Taylor, J. O. N. A. T. H. A. N. (s.f.). StatsModels: Statistics in Python — statsmodels 0.9.0 documentation. Recuperado 20 abril, 2019, de <https://www.statsmodels.org/stable/index.html>

²² ICFES, Mejor Saber – Valor agregado y aporte relativo, ¿Qué es valor agregado (VA)? URL:

<http://www.icfes.gov.co/valor-agregado-y-aporte-relativo>

²³ Di Pierro, M. A. S. S. I. M. O., Mulone, M. A. R. T. I. N., & Maldonado, J. E. N. N. I. F. E. R. (s.f.). web2py - Introducción. Recuperado 21 abril, 2019, de <http://www.web2py.com/books/default/chapter/36/01/introduccion>

2.2 Marco Teórico

2.2.1 Modelo Lineal Jerárquico

Los modelos lineales jerárquicos son una técnica de regresión que toma en cuenta la estructura jerárquica de los datos.²⁴ Se han denominado modelos de coeficientes aleatorios, modelos lineales multinivel y modelos de componentes de covarianza entre otros,²⁵ pero el término de modelo lineal jerárquico captura dos características que definen a los modelos en dos niveles diferentes que fueron tomados en el proyecto de grado de la UTP descritos a continuación:

Nivel 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

Dónde:

Y_{ij} = puntaje del estudiante i de la Universidad j

β_{0j} = intercepto de la universidad j

β_{1j} = media del rendimiento de la universidad j

e_{ij} = residuos entorno del rendimiento de la universidad j

Nivel 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j}$$

Dónde:

β_{0j} = media del rendimiento de la universidad j

γ_{00} = media general de toda la muestra

u_{0j} = error aleatorio.²⁶

²⁴ Universidad Veracruzana – Laboratorio de investigación y Asesoría estadística, LINA-E- Revista de ciencias básicas, Volumen 4 número 2 (Diciembre 2006) pág. 20 – 28

²⁵ Stephen W. Raudenbush, Michigan State University, East Lansing – Hierarchical Linear Models and Experimental Design. URL: http://jakewestfall.org/misc/Raudenbush_1993.pdf

²⁶ Universidad Tecnológica de Pereira, Juan Carlos Barrero - Modelo de Valor Agregado para la carrera de Ingeniería de Sistemas y Computación a partir de Saber Pro 2015 y 2016, pág. 26 y 27

2.2.2 Los modelos lineales mixtos

El Dr. Zuur²⁷ explica que los modelos lineales mixtos, los cuales también pueden ser conocidos como modelos de efectos mixtos o modelos multinivel, son modelos utilizados principalmente cuando los datos poseen una estructura jerárquica o algún tipo de agrupación, como lo pueden ser los diseños de medidas repetida, las series temporales, diseños anidados, entre otros.

Correa²⁸ explica que los modelos mixtos son una herramienta útil para el análisis estadístico. Estos abarcan los modelos lineales tradicionales, así como sus subunidades dentro del análisis estadístico. Los modelos lineales mixtos se caracterizan por su flexibilidad, ya que permiten obtener un perfil individual independiente del perfil promedio estadístico estudiado.

Según lo mencionado por Pinheiro²⁹ explica que los modelos lineales mixtos poseen dos efectos, los efectos fijos (efectos de interés para el experimentador) y los efectos aleatorios (poseen coeficientes cuyos niveles son una realización de los posibles niveles provenientes a una población). Para poder obtener un mejor resultado de estos dos componentes, es vital la selección de la mejor estructura, ya que si la estructura de efectos aleatorios se encuentra mal definida por consiguiente afectara los resultados de estimación de los efectos fijos y es importante considerar que en la mayoría de casos se hace en mayor énfasis en los resultados y datos provenientes de los efectos fijos.

2.2.3 Criterio de información de Akaike (Experian Commercial Risk Database AIC)

Como objeto Márquez³⁰ AIC consiste en: “el proceso de construcción de un modelo para una serie temporal se basa en la utilización de un conjunto de tests de hipótesis que van a permitir ir definiendo paso a paso el modelo que mejor se ajusta”.

Es un modelo de fácil formulación y aplicación, generalmente es utilizado para la obtención de modelos de análisis factorial, análisis de varianza y regresión múltiple. Se encarga de comparar la complejidad de un modelo estimado con la forma en la que el modelo se ajusta a los datos. Este modelo se basa en la medida de información de Kullback-Leibler, la cual permite calcular la distancia entre dos distribuciones.

²⁷ Zuur, A., Ieno, E., & Smith, G. (2007). *Analysing Ecological Data* (p. 125). Dordrecht: Springer.

²⁸ Correa Morales, J., & Salazar Uribe, J. (2016). *Introducción a los modelos mixtos* (1st ed., pp. 11-241). Medellín, Colombia.

²⁹ Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2013). nlme: Linear and nonlinear mixed effects models. *R package version*, 3(1), 111.

³⁰ Márquez Cebrián, M. D. (2002). *Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación*. Universitat Politècnica de Catalunya.

El cálculo AIC se realiza por medio de la siguiente expresión³¹:

$$AIC = 2K - 2\ln(L)$$

Criterio de información Bayesiano (Bayesian Information Criterion BIC) fue desarrollado por Schwarz (1978). En el área de la estadística, el criterio de información bayesiano (BIC) es considerado como un criterio para la selección de modelos de un conjunto finito de modelos.³²

BIC ha sido ampliamente utilizado para la identificación de modelos en series de tiempo y regresión lineal. Puede, sin embargo, aplicable a cualquier conjunto de modelos basados en máxima verosimilitud.

El criterio de información bayesiana (BIC) se define como³³:

$$BIC = -2 \cdot \ln(L) + k \ln(n)$$

n = tamaño de la muestra

k = el número de parámetros libres a estimar

L = el valor maximizado de la función de probabilidad para el modelo estimado

³¹ Márquez Cebrián, M. D. (2002). *Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación*. Universitat Politècnica de Catalunya.

³² Schwarz, Gideon E. (1978). "Estimating the dimension of a model". *Annals of Statistics* 6 (2): 461–464. doi:10.1214/aos/1176344136. MR468014.

³³ Calderón Rivera, D., Navarrete López, C., & Díaz Arévalo, J. (2016). Ajustes de distribuciones probabilísticas para la variable temperatura media multianual para el departamento de Boyacá (Colombia). *Ingeniería Y Región*, 14(2), 113-124.

3. ESTADO DEL ARTE

La magister en ciencias económicas Issabella Muñoz³⁴ plantea un modelo lineal jerárquico para calcular el valor agregado de cada universidad y tener así una clasificación de las universidades por características. Esto se realiza evaluando al individuo justo antes de entrar a la universidad y posteriormente al finalizar la misma. Los resultados obtenidos de esta investigación muestran que existen universidades que generan un mayor valor agregado que otras con respecto a la parte cognitiva de los alumnos. Los datos usados por la autora son los componentes de lectura crítica de las pruebas Saber 11 y Saber Pro del año 2014.

Por otro lado, Isáziga et al.³⁵, realizan un estudio el cual presenta una propuesta metodológica que permite estimar la diferencia entre los desempeños relativos de los estudiantes antes y después de un proceso formativo universitario, tomando como referente el Examen para el Ingreso a la Educación Superior (Saber 11) y el Examen de Estado de Calidad de la Educación Superior (SaberPRO); haciendo un análisis comparativo por Consejos Regionales de Planificación (Corpes). La metodología que se propuso calcula un porcentaje de efectividad máxima (Prueba de Wilcoxon) que permite analizar datos de diseños con medidas repetidas sobre las variables tomadas de las bases de datos de las pruebas Saber11 y Saber PRO, aplicadas en el periodo 2000-2009. Estas variables son: "documento de identidad", "nombre", "institución educativa", "departamento y región Corpes".

En la tesis de Ramiro Rodríguez Revilla de la Universidad Santo Tomás denominada Medición del valor agregado para la educación superior en Bogotá³⁶; Ramiro presenta un modelo lineal jerárquico que mide el valor agregado en las instituciones de educación superior de Bogotá utilizando las bases de datos de las pruebas de estado Saber 11 para el periodo 2002-1 – 2008-2 y Saber Pro para el periodo 2007-1 – 2012-2 facilitadas por el Instituto Colombiano para la Evaluación de la Educación y el Departamento Administrativo Nacional de Estadística, aislando los conocimientos previos adquiridos por los alumnos en la educación básica y media y los factores familiares y socioeconómicos de los mismos, de tal manera calculan una tabla en la cual se encuentran clasificadas diferentes instituciones de educación superior.

³⁴ Modelo de Valor Agregado: una implementación para el caso de la educación superior en Colombia, Issabella Muñoz Gómez, Universidad Nacional de Colombia, Facultad de Ciencias Económicas, Maestría en Ciencias Económicas, 2016

³⁵ La intervención académica en la construcción de una sociedad con calidad: análisis del valor agregado en el proceso formativo colombiano, Carlos-Hernán Isáziga-David, Jesús Gabalán-Coello, Fredy-Eduardo Vásquez-Rizo, HALLAZGOS / ISSN: 1794-3841 / Año 11, N.º 22 / Bogotá, D. C. / Universidad Santo Tomás / pp. 359-384, 2014.

³⁶ Medición del valor agregado para la educación superior en Bogotá, Ramiro Rodríguez Revilla, Universidad Santo Tomás, Facultad de Ciencias Económicas y Administrativas, Maestría en Ciencias Económicas, 2015.

4. METODOLOGÍA

4.1 Metodología de Desarrollo

En este proyecto de investigación se realizó un análisis de qué metodología aplicar, entre ellas, se analizaron metodologías tradicionales y metodologías ágiles, teniendo en cuenta estas últimas, ya que ofrecen una solución a medida del software y que una de las cualidades más destacables es su sencillez tanto en su aprendizaje como en su aplicación, reduciendo así mismo los costos de implantación en un equipo de desarrollo.³⁷ Las metodologías ágiles son flexibles, pueden ser modificadas para que se ajusten a la realidad del ciclo del proyecto que se va a desarrollar³⁸.

Para definir la metodología más práctica a utilizar tomamos en cuenta el alcance que va a tener el proyecto y los modelos de prototipo que puedan asegurar el correcto funcionamiento, muestre los resultados deseados y sea un acuerdo mutuo entre el investigador y los desarrolladores del software.

Teniendo los resultados del análisis, determinamos que la metodología **Rapid Application Development (RAD)** como metodología ágil se puede aplicar en este caso, considerando que el tiempo es reducido ya que *James Martin* quien argumentó que “*los desarrolladores pueden lograr grandes reducciones en tiempo y costo, entregando sistemas de información de gran calidad*”³⁹, por tal motivo esta metodología se ajusta al proyecto y como tal se presentará en el transcurso de un periodo académico.

La metodología **RAD** está centrada principalmente en ejecutar los procesos como su nombre lo indica, de una manera rápida. Combina el uso de técnicas de prototipado evolutivo y time boxing y puede desglosar el proceso de varias maneras, pero en general sigue cuatro fases principales.⁴⁰

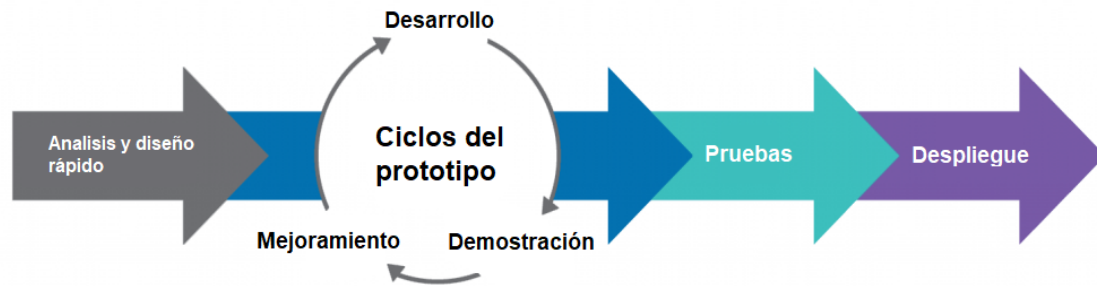
³⁷ Universidad Politécnica de Valencia José H. Canós, Patricio Letelier y M^a Carmen Penadés - Metodologías Ágiles en el Desarrollo de Software. 2012. URL: <http://roa.ult.edu.cu/bitstream/123456789/476/1/TodoAgil.pdf>

³⁸ Universidad Autónoma del Caribe, Revisión de metodologías ágiles para el desarrollo de software. PROSPECTIVA, vol. 11, núm. 2, julio-diciembre, 2013, pág. 30-39. URL: <https://www.redalyc.org/pdf/4962/496250736004.pdf>

³⁹ James Martin, (1991) Rapid Application Development; Macmillan Publishing Co., Inc.

⁴⁰ What is Rapid Application Development (RAD) - CASEMaker Inc. URL: http://www.iro.umontreal.ca/~dift6803/Transparents/Chapitre1/Documents/rad_wp.pdf

Ilustración 1 - Methodology RAD.



Fuente: <https://www.bairesdev.com/about/methodology/>.

Planificación de los requisitos: En esta fase se determina el alcance del proyecto, donde los desarrolladores e implicados pautan las metas y expectativas a alcanzar, así como los problemas actuales y potenciales que se deben abordar durante la construcción.

Diseño: Se construye el diseño del usuario a partir de varias iteraciones de prototipos, esta fase se diferencia de otras metodologías de desarrollo ya que el usuario final trabaja mano a mano con los desarrolladores para garantizar que todas las necesidades se cumplan en cada paso del proceso del diseño.

Construcción y desarrollo: Se toman los prototipos y los sistemas de prueba diseñados en la fase anterior y se convierten en un modelo de trabajo. Esta fase se descompone en varios pasos más pequeños:

- Preparación para la construcción rápida.
- Desarrollo de programas y aplicaciones.
- Codificación
- Unidad, integración y pruebas del sistema.

Transición: Como última fase, la implementación del producto terminado, este incluye la conversión de datos, las pruebas y el cambio del nuevo sistema, así como la capacitación del usuario.

5. REQUERIMIENTOS

5.1 Especificación de los requerimientos

A partir de la información entregada por la Universidad Católica de Colombia, con la metodología que están usando para el cálculo del valor agregado usando modelos lineales jerárquicos, se realizó la especificación de requerimientos según el estándar IEEE 830 (ver anexo A).

6. DISEÑO

6.1 Normalización de Base de Datos

Se realizó la normalización de la base de datos aplicando las tres primeras formas de la siguiente manera:

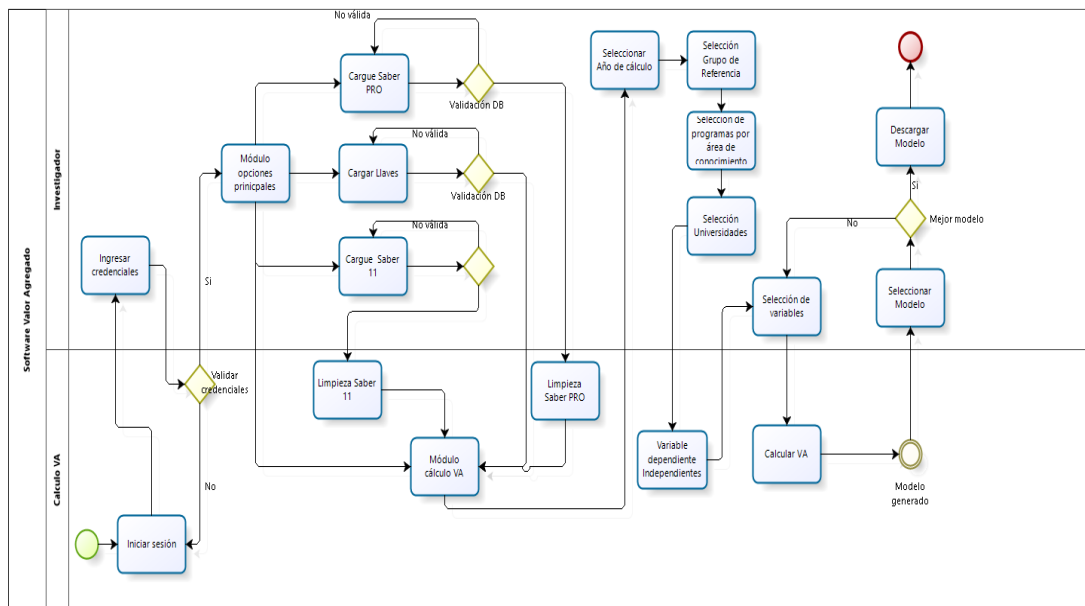
En la primera forma normal, las columnas de todas las tablas se transformaron en un valor único o atómico para ser combinadas en una sola estructura.

Para la segunda forma normal, se mantienen separadas las bases Saber Pro y las bases Saber 11 de forma organizada. La unión de estas dos se hace por medio de la tabla “Llaves de cruce” la cual contiene la llave de la prueba Saber 11 presentada por el estudiante al terminar sus estudios de bachillerato y la llave que corresponde al mismo estudiante en la prueba saber Pro.

Teniendo las tablas en la 2FN, sin dependencias transitivas y dependiendo solamente de la clave primaria, podemos dar por normalizada en la tercera forma la base de datos.

6.2 Diagrama de ejecución

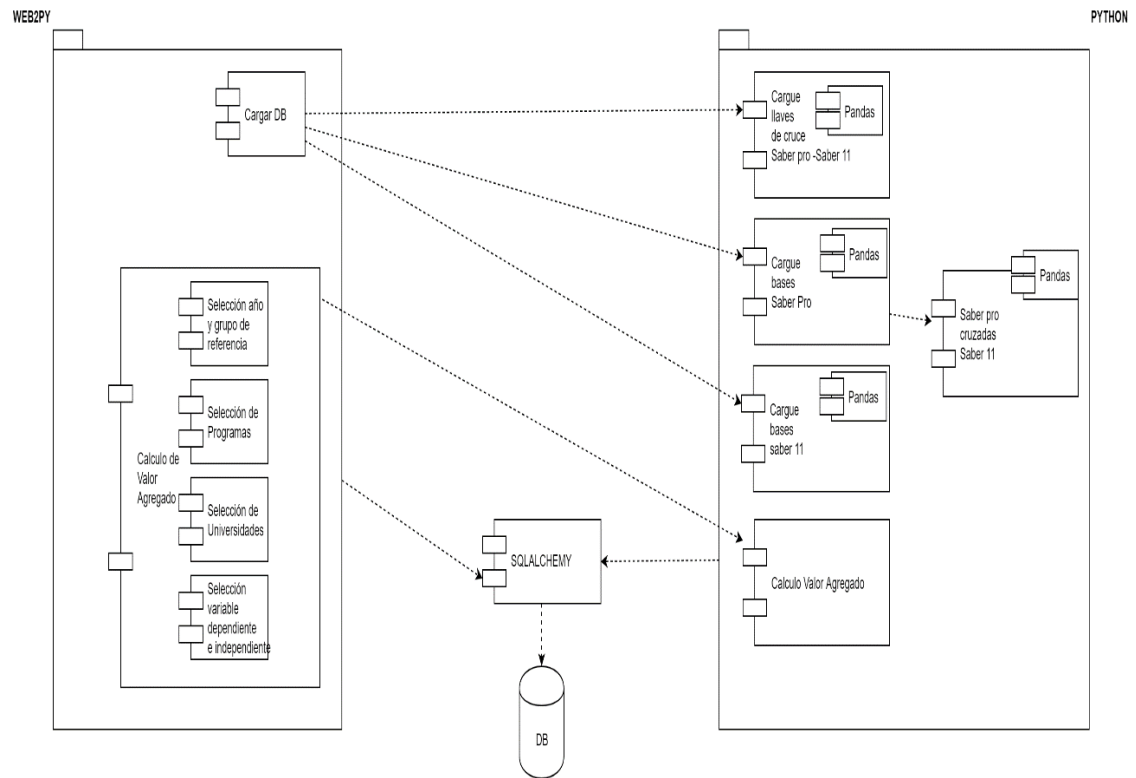
Ilustración 2 - Diagrama de ejecución del software.



Fuente: Los autores.

6.3 Diagrama de Componentes

Ilustración 3 - Diagrama de Componentes.

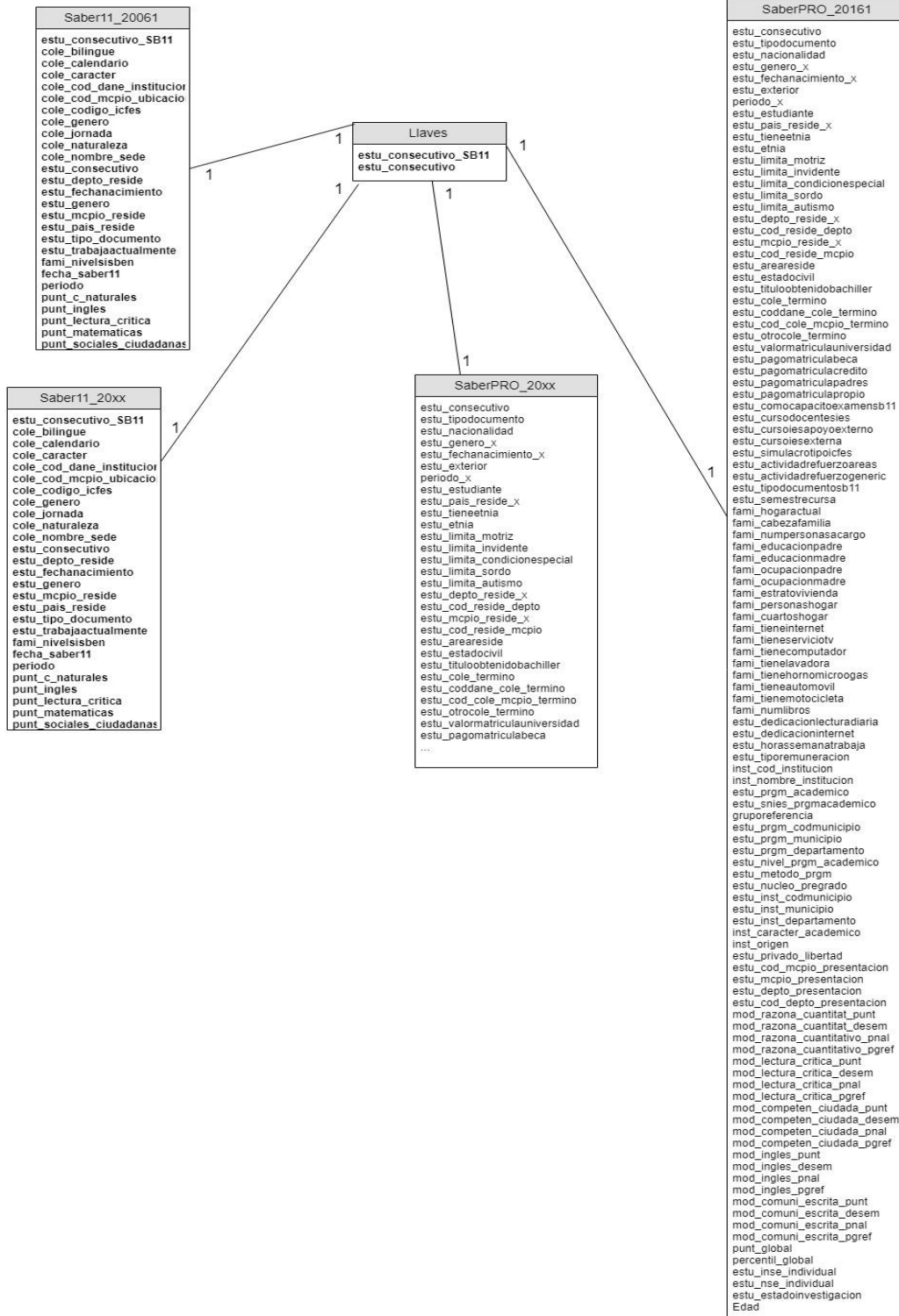


Fuente: Los autores.

6.4 Modelo Entidad Relación

Ilustración 4 - Modelo Entidad Relación.

Modelo ER

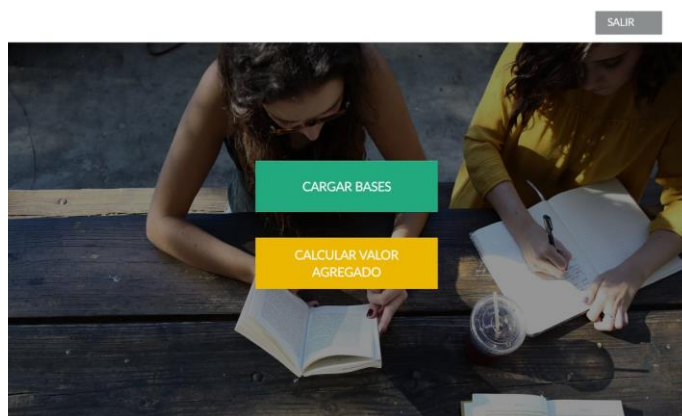


Fuente: Los autores.

6.5 Diseño de los MockUps

Según como se muestra en la metodología aplicada para este proyecto se realizaron prototipos y diseños, de la mano con el investigador o usuario final para garantizar que la solicitud y requerimientos responderán a cabalidad con las exigencias a cumplir en el desarrollo. Se pudo ajustar el modelo a medida que se avanzó hasta lograr un diseño satisfactorio que se muestra en los siguientes mockups:

Ilustración 5 - Pantalla inicial - Calculo Valor Agregado.



Fuente: Los autores.

Pantalla principal para acceso al módulo del cálculo de valor agregado (Ilustración 5). En esta pantalla se puede elegir si el usuario desea cargar las bases de las pruebas Saber Pro, Saber 11 y las llaves de cruce entre las bases, si ya están totalmente cargadas puede elegir la opción de calcular el valor agregado.

Ilustración 6 - Pantalla Cargar Bases de Datos.



Fuente: Los autores.

Para cargar las bases de datos el investigador debe seleccionar la opción de base que desea actualizar (Ilustración 6), esta muestra tres pestañas diferentes “Saber 11”, “Saber Pro” y “Llaves” ya que anualmente son presentadas por los estudiantes y deben ser actualizadas en el software de cálculo de Valor Agregado.

Ilustración 7 - Pantalla selección año y grupo de referencia ICFES.



Fuente: Los autores.

Si selecciona la opción de calcular Valor Agregado, en la pantalla se muestra dos opciones que permiten seleccionar el año a calcular el valor agregado de las pruebas Saber Pro (Ilustración 7), y la otra opción permite escoger los grupos de referencia por áreas del conocimiento del ICFES.

Ilustración 8 - Pantalla Selección programas según grupo de referencia ICFES.



Fuente: Los autores.

El investigador escoge las carreras que desea comparar después de ya haber seleccionado el grupo de referencia (Ilustración 8), los checklist que seleccione son a los que se les aplica el cálculo de valor agregado.

Ilustración 9 - Pantalla Selección Universidades según programas escogidos.



Fuente: Los autores.

En la pantalla de elección de universidades según programas escogidos (Ilustración 9), se muestran las universidades seleccionadas anteriormente, dando a conocer cuál fue la cantidad de estudiantes que presentaron las pruebas Saber en el año seleccionado.

Ilustración 10 - Pantalla Cálculo de Valor Agregado.



Fuente: Los autores.

La última pantalla (Ilustración 10), muestra las variables independientes que se pueden seleccionar y de ellas se genera el cálculo del valor agregado, se muestra el modelo actual y el modelo final, dando así la posibilidad de que el investigador tome opción el cual considere es el mejor modelo para al final mostrar los resultados del cálculo generado.

7. DESARROLLO

Para el desarrollo del software, en la parte de código backend, se utilizó el lenguaje de programación Python en su versión 3.5.3 junto con las bibliotecas de pandas, sqlalchemy y statsmodel. Todas estas librerías se encuentran contenidas en un virtual environment (entorno aislado de python) creado con la herramienta Virtualenv en su versión 16.4.3.

Para el desarrollo de la interfaz gráfica (front) se utilizó el framework Web2Py, este framework utiliza JQuery. La parte visual es un template de Bootstrap.

Para la persistencia de datos se utilizó el gestor de base de datos MariaDB en su versión 10.1.37.

7.1 Limpieza de archivos

En la limpieza de los archivos de las pruebas Saber 11, Saber Pro y las Llaves de cruce entre saber 11 y saber pro se realiza una lectura del archivo txt y se realizan los siguientes reemplazos:

Tabla 1 - Limpieza de caracteres.

CARÁCTER A REEMPLAZAR	CARÁCTER DE REEMPLAZO
á	a
é	e
í	i
ó	o
ú	u
ñ	nh
Á	A
É	E
Í	I
Ó	O
Ú	U
Ñ	NH
“	”
”	”
–	-
Ã	<vacío>

o (<0x82>)	<vacío>
-	<vacío>
Â	<vacío>
»	<vacío>
#	<vacío>
¿	<vacío>
;	nh
¬	~
	~
ï	~

Fuente: Los autores.

Estos caracteres son reemplazados para estandarizar los datos de las diferentes columnas de las bases y así mantener la similitud de los datos cualitativos entre un registro y otro, también para que la codificación de ellos no genere errores al momento de realizar tanto el cargue en la base de datos como la ejecución de los siguientes procesos. Luego de que los archivos son limpiados, se almacenan en un nuevo archivo cuyo nombre será el original del archivo sin extensión, más un guion bajo seguido de la palabra “clean” (nombrearchivo_clean). Si la base Saber 11 ya se encuentra cargada, se omite la limpieza y no se hace ningún cargue a la base de datos.

7.2 Cargue de archivos a la base de datos

7.2.1 Cargue de Pruebas Saber 11

Una vez generados los archivos cuyo nombre termina en “_clean” de las bases saber 11, el software procede a leer los archivos y genera un dataframe el cual es la principal estructura de datos de pandas; una estructura de datos tabular de dos dimensiones, mutable, potencialmente heterogénea con ejes etiquetados (filas y columnas). Puede considerarse como un contenedor tipo dict (diccionario de python) para objetos de la serie⁴¹, es prácticamente una tabla de una base de datos solo que cargada en memoria. A este dataframe de la base saber 11 se le realizar unos ajustes en las columnas de la siguiente manera:

⁴¹ pandas.DataFrame — pandas 0.23.4 documentation. (s.f.). Recuperado 22 abril, 2019, de <https://pandas.pydata.org/pandas-docs/version/0.23.4/generated/pandas.DataFrame.html>

Tabla 2 - Normalización de columnas.

COLUMNA	CAMBIO REALIZADO SOBRE ELLA
estu_fechanacimiento	si la columna no existe, se concatenan las columnas estu_nacimiento_anno, estu_nacimiento_mes y estu_nacimiento_dia de tal forma que conformen la fecha de nacimiento de la persona en formato YYYY-MM-DD ya que en las pruebas del periodo 2012-2 para atras no existia esta columna.
punt_c_naturales	si la columna no existe esta se creada a partir del promedio de la sumatoria entre los puntajes de punt_biologia, punt_fisica y punt_quimica, ya que apartir del periodo 2014-2, desaparecieron de la prueba saber 11 las anteriores columnas y quedó punt_c_naturales.
punt_lectura_critica	si la columna no existe se renombra la columna punt_lenguaje. Este renombramiento se hace a las bases de los periodos anteriores al 2014-1
punt_ingles	si la columna no existe se renombra la columna punt_idioma. Este renombramiento se hace a las bases de los periodos 2006-1 y 2006-2
estu_trabajaactualmente	si la columna no existe se renombra la columna estu_trabaja y si esta última tampoco existe se crea la columna con el contenido vacío.
fami_nivelsisben	si la columna no existe se renombra la columna fami_nivel_sisben y si esta última tampoco existe se crea la columna con el contenido vacío.
estu_depto_reside	si la columna no existe se renombra la columna estu_reside_depto. Este renombramiento se hace a las bases de los periodos anteriores a 2013-2
estu_mcpio_reside	si la columna no existe se renombra la

	columna estu_reside_mcpio. Este renombramiento se hace a las bases de los periodos anteriores a 2013-2
cole_cod_dane_institucion	si la columna no existe se renombra la columna cole_cod_dane_establecimiento. Este renombramiento se hace a las bases de los periodos posteriores a 2012-2
cole_codigo_icfes	si la columna no existe se renombra la columna cole_cod_icfes. Este renombramiento se hace a las bases de los periodos anteriores a 2013-2
estu_tipo_documento	si la columna no existe se renombra la columna estu_tipodocumento. Este renombramiento se hace a las bases de los periodos posteriores a 2012-2
punt_sociales_ciudadanas	si la columna no existe se renombra la columna punt_ciencias_sociales. Este renombramiento se hace a las bases de los periodos anteriores a 2013-2

Fuente: Los autores.

Este ajuste se hace, ya que las bases de Saber 11 con el transcurso de los años cambian el nombre de las columnas y los datos. Para la conformación de la base final se requiere que las columnas posean el mismo nombre y la misma cantidad de columnas. Una vez hecho esto, se procede a cargar la información a la base de datos desde el dataframe y solamente con las siguientes columnas:

Tabla 3 - Columnas finales Saber 11.

cole_bilingue	cole_calendario	cole_caracter	cole_cod_dane_institucion
cole_codigo_icfes	cole_cod_mcpio_ubicacion	cole_genero	cole_jornada
cole_naturaleza	cole_nombre_sede	estu_consecutivo	estu_genero
estu_fechanacimiento	estu_pais_reside	estu_depto_reside	estu_mcpio_reside
estu_tipo_documento	periodo	punt_c_naturales	punt_ingles
punt_matematicas	punt_sociales_ciudadanas	punt_lectura_critica	estu_trabajaactualmente
fami_nivelsiben	fecha_saber11		

Fuente: Los autores.

Pandas permite enviar el dataframe directamente a una base de datos y así generar la tabla correspondiente en MySQL.

7.2.2 Cargue de Llaves de cruce entre Saber 11 y Saber pro

Una vez generado el archivo “_clean” de las llaves, el software procede a leer en un dataframe el archivo y se carga sin ninguna modificación a la base de datos, estas llaves son utilizadas para poder realizar la unión de información entre las pruebas Saber 11 y las pruebas Saber Pro, es decir, tomar la información de la prueba Saber Pro de una persona y enlazar la información correspondiente de la prueba Saber 11 de la misma persona, creando así un registro con toda la información necesaria para realizar el cálculo del valor agregado.

7.2.3 Cargue de Pruebas Saber Pro

Una vez generado el archivo _clean de las pruebas Saber Pro el software procede a cargar en un dataframe el contenido del archivo y se realizan los siguientes cambios a los valores de las siguientes variables como se muestran en las tablas:

- Nivel educativo de la madre y del padre.

Tabla 4 - Nivel educativo padres – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
Secundaria completa	11
Secundaria incompleta	9
Educación profesional completa	16
Educación profesional incompleta	14
Técnica o tecnológica completa	13
Técnica o tecnológica incompleta	11
Primaria completa	5
Primaria incompleta	3
Postgrado	18

Fuente: Los autores.

- Horas de trabajo.

Tabla 5 - Horas trabajo de estudiante – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
Menos de 10 horas	5

Mas de 30 horas	40
Entre 21 y 30 horas	25
Entre 11 y 20 horas	15

Fuente: Los autores.

- Estrato socio económico.

Tabla 6 - Estrato Socioeconómico – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
Estrato 0	0
Estrato 1	1
Estrato 2	2
Estrato 3	3
Estrato 4	4
Estrato 5	5
Estrato 6	6

Fuente: Los autores.

- Género

Tabla 7 - Género – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
M	1
F	0

Fuente: Los autores.

- Tipo de Universidad:

Tabla 8 - Tipo Universidad – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
NO OFICIAL - CORPORACION	1
NO OFICIAL - FUNDACION	1
OFICIAL MUNICIPAL	0
OFICIAL DEPARTAMENTAL	0
OFICIAL NACIONAL	0
REGIMEN ESPECIAL	0

Fuente: Los autores.

- Metodología de estudio:

Tabla 9 - Metodología de estudio – Valor.

VALOR CUALITATIVO	VALOR CUANTITATIVO
PRESENCIAL	1
DISTANCIA	0
DISTANCIA VITUAL	2
SEMI-PRESENCIAL	3

Fuente: Los autores.

- Valor de la matrícula.

Tabla 10 - Valor promedio de matrícula.

VALOR CUALITATIVO	VALOR CUANTITATIVO
Entre 500 mil y menos de 1 millón	750000
Entre 1 millón y menos de 2.5 millones	1750000
Entre 2.5 millones y menos de 4 millones	3250000
Menos de 500 mil	250000
Mas de 7 millones	8000000
Entre 4 millones y menos de 5.5 millones	4750000
Entre 5.5 millones y menos de 7 millones	6250000
No pago matricula	0

Fuente: Los autores.

Este dataframe es cargo en la base de datos generando una tabla de las pruebas Saber Pro del año cargado.

7.2.4 Generación de tabla Saber Pro Agregado saber 11

Una vez cargada en la base de datos la prueba Saber Pro, el software mantiene en memoria el dataframe y seguido a esto se realiza una consulta a la base de datos para extraer en un dataframe las Llaves de cruce entre Saber 11 y Saber pro, Luego se recorren cada una de las bases de datos Saber 11, extrayendo la información de estas a memoria y realizando un Merge entre el dataframe de las pruebas Saber Pro, Saber 11 y Las llaves de cruce (el Merge realiza la misma función que un join en una base de datos solo que en memoria) donde las llaves de intersección entre los datos son las llaves entregadas por el icfes y la llave correspondiente en las bases Saber Pro y Saber 11. Este Merge retorna un

dataframe con toda la información y este se carga en la base de datos para posteriormente ser utilizado en la extracción del grupo de referencia creado por el usuario y consecuente a ello el cálculo del valor agregado.

7.3 Grupo de referencia

Para crear el grupo de referencia se requiere definir ciertos filtros para extraer solamente la información necesaria de la base de datos. Estos filtros son:

7.3.1 Elección de año sobre la cual se va a realizar el cálculo

En la página web en el módulo de Cálculo de Valor Agregado, el usuario escoge la base de datos del año correspondiente sobre la cual va a trabajar, esta se encuentra en una lista desplegable la cual solo permite la selección de un año, este año corresponden a las bases saber pro cargadas anteriormente.

7.3.2 Elección de grupos de referencia por áreas de conocimiento del ICFES

En las bases de las pruebas Saber Pro, el ICFES agrega una columna en la cual especifica la persona a que área de conocimiento pertenece, las áreas de conocimiento para el ICFES son bellas artes y diseño, ciencias naturales y exactas, ciencias sociales, humanidades, derecho, comunicación periodismo y publicidad, ciencias militares y navales, ciencias agropecuarias, administración y afines, educación, arquitectura y urbanismo, ingeniería, salud, medicina, recreación y deportes, economía, contaduría y afines, psicología, enfermería⁴². Una vez escogido el año sobre el cual se va a trabajar, se habilita la opción de escoger un área de conocimiento sobre la cual va a trabajar en una lista desplegable dentro del módulo del cálculo de valor agregado.

7.3.3 Elección de programas según grupo de referencia por área de conocimiento del ICFES escogido

En las bases de las pruebas Saber Pro, el ICFES agrega una columna en la cual especifica la persona a qué programa pertenece, los programas para el ICFES son:

Tabla 11 - Programas Académicos según Grupo de Referencia ICFES.

Consecutivo	Nombre del Grupo de Referencia	Nivel	Programas académicos
REF1		UNIVERSITARIA	ARTE

⁴² MIN EDUCACIÓN. (s.f.). Grupos de referencia Saber Pro 2017. Recuperado 24 abril, 2019, de <http://www2.icfesinteractivo.gov.co/investigacionFormulario/terminos-de-uso/item/2295>

	BELLAS ARTES Y DISEÑO	ARTE DANZARIO ARTE DRAMÁTICO ARTES AUDIOVISUALES ARTES CULINARIAS Y GASTRONOMÍA ARTES DE LA GRABACIÓN Y PRODUCCIÓN MUSICAL ARTES PLÁSTICAS BELLAS ARTES CINE Y TELEVISIÓN COMUNICACIÓN VISUAL Y MULTIMEDIA CONSERVACIÓN Y RESTAURACIÓN CULINARIA Y GASTRONOMÍA DANZA DIRECCIÓN DE BANDA DISEÑO DISEÑO DE MODA DISEÑO GRÁFICO DISEÑO INDUSTRIAL DISEÑO INTERIOR GASTRONOMÍA MUSEOLOGÍA MÚSICA AFINES
REF2	CIENCIAS NATURALES Y EXACTAS	UNIVERSITARIA ASTRONOMÍA BIOLOGÍA BIOQUÍMICA ECOLOGÍA ESTADÍSTICA FÍSICA GEO CIENCIAS GEOLOGÍA INFORMÁTICA MATEMÁTICA

			INGENIERÍA BIOLÓGICA
			INGENIERÍA FÍSICA
			MATEMÁTICAS
			MICROBIOLOGÍA
			OCEANOGRAFÍA
			QUÍMICA
			AFINES
REF3	CIENCIAS SOCIALES	UNIVERSITARIA	ANTROPOLOGÍA
			ARQUEOLOGÍA
			CIENCIA POLÍTICA
			CIENCIAS SOCIALES
			DESARROLLO FAMILIAR
			ESTUDIOS Y GESTIÓN CULTURAL
			GEOGRAFÍA
			GESTIÓN CULTURAL
			HISTORIA
			RELACIONES INTERNACIONALES
			RELACIONES INTERNACIONALES Y ESTUDIOS POLÍTICOS
			SISTEMAS DE INFORMACIÓN Y DOCUMENTACIÓN
			SOCIOLOGÍA
			TERAPIAS PSICOSOCIALES
			TRABAJO SOCIAL
			AFINES
REF4	HUMANIDADES	UNIVERSITARIA	ARCHIVÍSTICA
			CIENCIAS BÍBLICAS
			CREACIÓN LITERARIA
			ESTUDIOS LITERARIOS
			FILOSOFÍA
			HISTORIA
			LENGUAS MODERNAS
			LITERATURA

			TEOLOGÍA
			TRADUCCIÓN
			TRADUCCIÓN SIMULTANEA
			AFINES
REF5	DERECHO	UNIVERSITARIA	CRIMINALÍSTICA
			DERECHO
			DERECHO Y CIENCIAS POLÍTICAS
			GOBIERNO
			INVESTIGACIÓN CRIMINAL
			JURISPRUDENCIA
			LEYES Y JURISPRUDENCIA
			AFINES
REF6	COMUNICACIÓN, PERIODISMO Y PUBLICIDAD	UNIVERSITARIA	BIBLIOTECOLOGÍA
			COMUNICACIÓN AUDIOVISUAL Y DIGITAL
			COMUNICACIÓN SOCIAL Y PERIODISMO
			COMUNICACIÓN GRAFICA PUBLICITARIA
			COMUNICACIÓN ORGANIZACIONAL
			COMUNICACIÓN PUBLICITARIA
			COMUNICACIÓN SOCIAL Y PERIODISMO
			MERCADEO PUBLICIDAD Y VENTAS
			MERCADEO Y PUBLICIDAD
			PERIODISMO
			PUBLICIDAD
			AFINES
REF7	CIENCIAS MILITARES Y NAVALES	UNIVERSITARIA	CIENCIAS MILITARES
			CIENCIAS MILITARES AERONÁUTICAS

			CIENCIAS NAVALES
			EDUCACIÓN FÍSICA MILITAR
			GERENCIA DE LA SEGURIDAD Y ANÁLISIS SOCIO POLÍTICO
			AFINES
REF8	CIENCIAS AGROPECUARIAS	UNIVERSITARIA	ACUICULTURA
			ADMINISTRACIÓN AGROINDUSTRIAL
			AGRONOMÍA
			INDUSTRIAS PECUARIAS
			INGENIERÍA AGRONÓMICA
			INGENIERÍA AGROPECUARIA
			MANEJO AGROECOLÓGICO Y DE POST-COSECHA
			MANEJO AGROFORESTAL
			MEDICINA VETERINARIA
			MEDICINA VETERINARIA Y ZOOTECNIA
			MERCADEO AGROINDUSTRIAL
			MERCADEO AGROPECUARIO
			ZOOTECNIA
			AFINES
REF9	ADMINISTRACIÓN Y AFINES	UNIVERSITARIA	ADMINISTRACIÓN DE EMPRESAS
			ADMINISTRACIÓN DE NEGOCIOS
			ADMINISTRACIÓN FINANCIERA
			ADMINISTRACIÓN COMERCIAL
			ADMINISTRACIÓN DE COMERCIO EXTERIOR

		ADMINISTRACIÓN DE CONSTRUCCIONES
		ADMINISTRACIÓN DE GESTIÓN HUMANA
		ADMINISTRACIÓN DE HOTELERÍA Y TURISMO
		ADMINISTRACIÓN DE OBRAS CIVILES
		ADMINISTRACIÓN DE SERVICIOS
		ADMINISTRACIÓN DEL MEDIO AMBIENTE
		ADMINISTRACIÓN EN SALUD
		ADMINISTRACIÓN FINANCIERA
		ADMINISTRACIÓN INDUSTRIAL
		ADMINISTRACIÓN LOGÍSTICA
		ADMINISTRACIÓN MARÍTIMA
		ADMINISTRACIÓN PÚBLICA
		BANCA Y COMERCIO EXTERIOR
		BANCA Y FINANZAS INTERNACIONALES
		COMERCIO EXTERIOR
		COMERCIO INTERNACIONAL
		CONTADURÍA
		CONTADURÍA PÚBLICA
		DIRECCIÓN Y ADMINISTRACIÓN DE EMPRESAS
		FINANZAS
		GERENCIA DE MERCADEO

			GESTIÓN EMPRESARIAL
			HOTELERÍA Y TURISMO
			LOGÍSTICA
			MARKETING
			MERCADEO
			RELACIONES ECONÓMICAS INTERNACIONALES
			TURISMO
			AFINES
REF10	EDUCACIÓN	UNIVERSITARIA	EDUCACIÓN BÁSICA
			FILOLOGÍA E IDIOMAS
			LICENCIATURA EN PEDAGOGÍA INFANTIL
			LICENCIATURA EN CIENCIAS
			LICENCIATURA EN EDUCACIÓN
			LICENCIATURA EN MATEMÁTICAS
			LICENCIATURA EN ARTE
			LICENCIATURA EN BÁSICA PRIMARIA
			LICENCIATURA EN BIOLOGÍA
			LICENCIATURA EN CIENCIAS NATURALES
			LICENCIATURA EN CIENCIAS SOCIALES
			LICENCIATURA EN EDUCACIÓN BILINGÜE
			LICENCIATURA EN ESPAÑOL Y COMUNICACIÓN
			LICENCIATURA EN FILOSOFÍA
			LICENCIATURA EN HUMANIDADES

			LICENCIATURA EN IDIOMAS EXTRANJEROS
			LICENCIATURA EN INFORMÁTICA
			LICENCIATURA EN MATEMÁTICAS
			LICENCIATURA EN MÚSICA
			AFINES
REF11	ARQUITECTURA Y URBANISMO	UNIVERSITARIA	ARQUITECTURA
			CONSTRUCCIÓN
			DISEÑO DE INTERIORES
			URBANISMO
			AFINES
REF12	INGENIERÍA	UNIVERSITARIA	ADMINISTRACIÓN AMBIENTAL
			ADMINISTRACIÓN DE INFORMÁTICA
			ADMINISTRACIÓN DE SISTEMAS DE INFORMACIÓN
			BIOINGENIERÍA
			BIOTECNOLOGÍA
			CONSTRUCCIONES CIVILES
			DISEÑO INDUSTRIAL
			INGENIERÍA AMBIENTAL
			INGENIERÍA BIOMÉDICA
			INGENIERÍA CIVIL
			INGENIERÍA COMERCIAL
			INGENIERÍA INDUSTRIAL
			INGENIERÍA MECÁNICA
			INGENIERÍA AGROFORESTAL
			INGENIERÍA AMBIENTAL
			INGENIERÍA CIVIL

			INGENIERÍA AGRONÓMICA
			INGENIERÍA BIOQUÍMICA
			INGENIERÍA DE MATERIALES
			INGENIERÍA DE MINAS
			INGENIERÍA DE PRODUCCIÓN
			INGENIERÍA DE SISTEMAS
			INGENIERÍA DE SOFTWARE
			INGENIERÍA ELÉCTRICA
			INGENIERÍA INFORMÁTICA
			INGENIERÍA QUÍMICA
			AFINES
REF13	SALUD	UNIVERSITARIA	BACTERIOLOGÍA Y LABORATORIO CLÍNICO
			FARMACIA
			FISIOTERAPIA
			FONOAUDILOGÍA
			GERONTOLOGÍA
			INSTRUMENTACIÓN QUIRÚRGICA
			MICROBIOLOGÍA Y BIOANÁLISIS
			NUTRICIÓN Y DIETÉTICA
			ODONTOLOGÍA
			OPTOMETRÍA
			QUÍMICA FARMACÉUTICA
			SALUD OCUPACIONAL
			TERAPIA FÍSICA
			TERAPIA OCUPACIONAL
AFINES			
REF14	MEDICINA	UNIVERSITARIA	MEDICINA

REF27	NORMALES SUPERIORES	NORMALISTA	CICLO COMPLEMENTARIO DE ESCUELAS NORMALES SUPERIORES
REF30	RECREACIÓN Y DEPORTES	UNIVERSITARIA	ACTIVIDAD FÍSICA Y DEPORTE
			CIENCIAS DEL DEPORTE Y LA RECREACIÓN
			CIENCIAS DEL DEPORTE Y DE LA EDUCACIÓN FÍSICA
			CULTURA FÍSICA Y DEPORTE
			DEPORTE
			ENTRENAMIENTO DEPORTIVO
			RECREACIÓN
			AFINES
REF31	ECONOMÍA	UNIVERSITARIA	ECONOMÍA
			ECONOMÍA Y FINANZAS INTERNACIONALES
			ECONOMÍA Y NEGOCIOS INTERNACIONALES
			FINANZAS Y COMERCIO INTERNACIONAL
			AFINES
REF32	CONTADURÍA Y AFINES	UNIVERSITARIA	CONTADURÍA PÚBLICA
			GOBIERNO Y ASUNTOS PÚBLICOS
			PROFESIONAL EN CONTADURÍA PÚBLICA
			AFINES
REF33	PSICOLOGÍA	UNIVERSITARIA	PSICOLOGÍA
			PSICOLOGÍA CON ÉNFASIS EN PSICOLOGÍA SOCIAL
			AFINES
REF34	ENFERMERÍA	UNIVERSITARIA	ENFERMERÍA

Fuente: <http://www2.icfesinteractivo.gov.co/investigacionFormulario/terminos-de-uso/item/1934-grupos-de-referencia-saber-pro-2018-1>.

Una vez seleccionada el área de conocimiento, se listan en un checklist los programas pertenecientes a esa área para que el usuario escoja uno o más programas para comparar⁴³.

7.3.4 Elección de universidades según programas escogidos.

Una vez seleccionados los programas, se listan los nombres de las universidades cuyos estudiantes presentaron la prueba Saber Pro y pertenecen a los programas seleccionados, la cantidad de estudiantes de la universidad que presentaron la prueba y también muestra en la página web, aquellas universidades cuya cantidad de estudiantes no son comparables con las demás universidades y son tomadas como datos atípicos, estos no entrarán dentro del cálculo de valor agregado.

7.3.5 Creación del grupo de referencia

Una vez escogido todos los anteriores filtros para el modelo, se conforma un grupo de referencia especializado, este con el fin de utilizar solamente las personas pertenecientes a la configuración escogida por el usuario y realizar el cálculo del valor agregado sobre esta información.

7.4 Cálculo del valor agregado

Una vez generado el grupo de referencia sobre el cual se va a realizar el cálculo de valor agregado, el usuario debe definir las variables con las cuales va a ejecutar el modelo. Estas variables son:

7.4.1 Elección de la variable dependiente

En la página web, en el módulo de Cálculo de Valor Agregado, el usuario escoge la variable dependiente con la cual va a realizar el cálculo, esta variable es una de las 5 competencias genéricas que el ICFES da en las pruebas Saber Pro, estas competencias son:

- Lectura Crítica.
- Razonamiento Cuantitativo.
- Competencias Ciudadanas.
- Comunicación Escrita.
- Inglés.

⁴³ MIN EDUCACIÓN. (s.f.). Grupos de referencia Saber Pro 2017. Recuperado 24 abril, 2019, de <http://www2.icfesinteractivo.gov.co/investigacionFormulario/terminos-de-uso/item/2295>

Esta es una variable requerida para el cálculo y se envía como parámetro al momento de realizar el mismo.

7.4.2 Elección de las variables independientes

Las variables independientes se muestran en la página web en un checklist, estas variables son:

- Puntaje en competencia genérica en saber 11.
- Edad.
- Nivel educativo de la madre y del padre.
- Horas de trabajo.
- Estrato socio económico.
- Género
- Tipo de Universidad:
- Metodología de estudio:
- Valor de la matrícula.

El usuario puede elegir una o varias variables independientes, esto con el fin de encontrar el modelo que más se ajuste a sus necesidades. Estas variables son requeridas para el cálculo y se envían como parámetro al momento de realizar el mismo.

7.4.3 Cálculo del valor agregado sobre el grupo de referencia

Una vez conformado el grupo de referencia y escogido la variable dependiente y las independientes, el usuario da click en el botón “calcular”. El software realiza el cálculo del valor agregado utilizando la función **mixedlm** de la librería **Statsmodels**:

```
mixedlm(var_dep+" ~ "+vars_indep, data, groups=data[sujeto])
```

Donde:

- **var_dep** = Variable dependiente (Competencias Genéricas)
- **vars_indep** = Variables independientes escogidas cuyo separador es el carácter ‘+’ (Edad+Género+Estrato+...+ni)
- **data** = Dataframe con la información extraída según los filtros del año, grupo de referencia del ICFES, programas y universidades.
- **groups** = Toma como parámetro la columna del nombre de las universidades (data[sujeto]).

Esta función retorna la siguiente tabla de resultado estadísticos para los parámetros escogidos:

Ilustración 11 - Resultado Modelo de regresión lineal Mixto.

Mixed Linear Model Regression Results						
Model:	MixedLM	Dependent Variable:	Razona_cuantitat			
No. Observations:	2897	Method:	REML			
No. Groups:	63	Scale:	479.7720			
Min. group size:	10	Likelihood:	-13100.7217			
Max. group size:	222	Converged:	Yes			
Mean group size:	46.0					
	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Intercept	96.887	5.527	17.529	0.000	86.054	107.721
Genero	8.611	1.015	8.483	0.000	6.622	10.600
Edad	-0.567	0.151	-3.745	0.000	-0.864	-0.270
matricula	0.729	0.675	1.081	0.280	-0.593	2.051
Estrato	1.188	0.507	2.341	0.019	0.193	2.182
trabajo	0.114	0.027	4.227	0.000	0.061	0.166
Educacionmadre	0.068	0.116	0.588	0.557	-0.160	0.297
Educacionpadre	0.192	0.100	1.925	0.054	-0.004	0.387
Matematicas	1.013	0.042	23.946	0.000	0.931	1.096
Metodologia	18.588	3.036	6.122	0.000	12.637	24.538
Oficial	-8.027	2.895	-2.773	0.006	-13.701	-2.353
Group Var	54.744	0.613				

Fuente: Los autores Generada por el Software de Valor Agregado.

El resultado presentado por el software se muestra en pantalla para que el usuario o investigador, analice los resultados y escoja el modelo más adecuado usando los criterios de significancia individual y los criterios de información. Al usuario se le presenta tanto el cálculo actual como el mejor cálculo según el requerimiento RF06.

Para poder calcular el Valor Agregado se utiliza la siguiente función:

```
predict = mdf.predict(data)
```

La cuál retorna los valores de los efectos fijo y aleatorios de cada registro de la información puesta en dicho modelo, el resultado de la diferencia de estos valores es el valor agregado.

7.5 Exportación de resultados

Una vez el usuario de clic en “Exportar resultados”, se generará un archivo en formato csv en el cual se encontrará el ranking de las universidades ordenado por el valor agregado que estas aportan a los estudiantes. La tabla exportada mantiene la siguiente estructura:

Tabla 12 - Ejemplo cálculo Valor Agregado.

inst_nombre_institucion	Cantidad Estu	Promedio Sab	Media Saber11	Promedio Saber	Media SaberPro	valor_agregado	Orden Saber	Orden Valor
UNIVERSIDAD DEL ATLANTICO-BI	115	53.49	54.05	170.54	168.81	6.23	11	1
COLEGIO MAYOR DE NUESTRA SE	205	60.67	54.05	187.45	168.81	5.25	3	2
UNIVERSIDAD AUTONOMA LATIN	135	52.19	54.05	170.56	168.81	5.13	10	3
UNIVERSIDAD EXTERNADO DE CC	203	61.94	54.05	188.69	168.81	4.52	2	4
UNIVERSIDAD DE MEDELLIN-MED	218	55.14	54.05	174.92	168.81	3.99	6	5
UNIVERSIDAD DE LOS ANDES-BO	174	63.95	54.05	190.04	168.81	3.51	1	6
UNIVERSIDAD LIBRE-PEREIRA	128	52.09	54.05	171.59	168.81	3.08	9	7
UNIVERSIDAD SANTO TOMAS-BU	138	56.34	54.05	175.39	168.81	2.06	5	8
UNIVERSIDAD SANTO TOMAS-BC	163	55.41	54.05	173.14	168.81	1.27	7	9
UNIVERSIDAD CATOLICA DE COL	207	52.33	54.05	166.0	168.81	0.72	13	10
UNIVERSIDAD LA GRAN COLOMB	281	50.9	54.05	163.84	168.81	0.58	14	11
UNIVERSIDAD MILITARNUEVA GF	194	55.62	54.05	170.51	168.81	0.49	12	12
PONTIFICIA UNIVERSIDAD JAVER	169	61.87	54.05	183.38	168.81	0.1	4	13
UNIVERSIDAD LIBRE-BOGOTA D.C	280	55.87	54.05	172.6	168.81	-1.32	8	14
CORPORACION UNIVERSITARIA C	117	48.66	54.05	157.31	168.81	-1.8	18	15
UNIVERSIDAD SANTIAGO DE CAL	125	48.97	54.05	157.34	168.81	-2.83	17	16
UNIVERSIDAD POPULAR DEL CES	131	52.37	54.05	157.37	168.81	-3.89	16	17
INSTITUCION UNIVERSITARIA CEI	121	49.21	54.05	154.03	168.81	-4.35	20	18
UNIVERSIDAD COOPERATIVA DE	425	49.01	54.05	154.18	168.81	-5.09	19	19
UNIVERSIDAD LIBRE-BARRANQUI	154	51.79	54.05	158.81	168.81	-8.45	15	20
UNIVERSIDAD SIMON BOLIVAR-E	229	47.32	54.05	147.27	168.81	-9.2	21	21

Fuente: Los autores – tabla exportada por el software.

8. PRUEBAS

8.1 Prueba de caja blanca:

Se realizaron las pruebas de caja blanca en el IDE “Spyder” ejecutando el backend en modo debug para probar cada uno de los componentes del sistema. Se hizo una prueba general de la conexión a la base de datos con el componente SQLAlchemy y se verificó que se pudiesen realizar consultas a la base de datos. La prueba sobre cada uno de los demás componentes fueron las siguientes:

8.1.1 Componentes front:

- **Módulo para el cargue de bases:**

Se realizó la prueba en la página web accediendo al cargue de cada base, comprobando que su visualización era la adecuada, no presentaba demoras y permitía subir los archivos correctamente.

- **Módulo para el cálculo de valor agregado:**

Se verificó el funcionamiento de este módulo ingresando a cada componente de este y rectificando sus funcionalidades las cuales son:

- **Selección del año y grupo de referencia:**

Se verificó que las listas desplegadas para la elección del año y grupo de referencia presentarán los datos correspondientes (años de las bases saber pro que están cargadas y los grupos de referencia para las diferentes carreras según el ICFES).

- **Selección de programas:**

Se verificó que el checklist de los programas correspondiera al año y grupo de referencia escogidos previamente y que permitiera realizar filtros sobre esa información.

- **Selección de universidades:**

Se verificó que el checklist de las universidades mostrará las universidades que cuentan o imparten los programas escogidos previamente y que permitiera realizar filtros sobre esa información.

- **Selección de variables dependientes e independientes:**

Se verificó que la lista desplegable de las variables dependientes sólo mostrará los 5 componentes genéricos evaluados en las pruebas Saber pro y que el checklist de las variables independientes tuviese únicamente las variables necesarias para el cálculo del valor agregado.

8.1.2 Componentes Backend

- **Pandas:**

Se comprobó este componente verificando la creación de los DataFrames y que el cargue en la base de datos de los mismos fuera correcta y completa.

- **Cargue de llaves Saber pro Saber 11:**

Se comprobó que el componente realizará la limpieza y cargue de las llaves de cruce de las pruebas Saber pro con las pruebas Saber 1 entregadas por el ICFES.

- **Cargue de bases Saber Pro:**

Se comprobó que el componente realizará la limpieza y cargue de las bases de resultados de las pruebas Saber pro entregadas por el ICFES.

- **Cargue de bases Saber 11:**

Se comprobó que el componente realizará la limpieza y cargue de las bases de resultados de las pruebas Saber 11 entregadas por el ICFES.

- **Creación tabla Saber Pro cruzado Saber 11:**

Se comprobó que el componente realizará la combinación de las bases Saber pro con las bases Saber 11 para conformar la tabla que utilizará el componente del cálculo de valor agregado para realizar el mismo.

- **Cálculo del Valor Agregado:**

Se comprobó que el componente calculará el valor agregado del grupo de referencia que crea el usuario en la web con los diferentes filtros que allí se le ofrecen y que exportará el ranking de las universidades una vez aplicado el modelo y ordenado de mayor a menor según el valor agregado que la universidad aporta a sus estudiantes.

8.2 Prueba de caja negra:

Se realizó la ejecución del software y se compararon los resultados de este con la ejecución del mismo modelo lineal mixto en el lenguaje R.

Ilustración 12 - Ejemplo de Modelo hecho en R Studio.

```

      AIC      BIC    logLik deviance df.resid
26214.4 26292.0 -13094.2  26188.4    2883

Scaled residuals:
   Min       1Q   Median       3Q      Max
-4.6439 -0.6526  0.0379  0.6501  5.4780

Random effects:
   Groups                Name      Variance Std.Dev.
PROinst_nombre_institucion (Intercept)  51.24    7.158
Residual                          478.60   21.877
Number of obs: 2896, groups: PROinst_nombre_institucion, 63

Fixed effects:
              Estimate Std. Error t value
(Intercept)  96.69627    5.51433   17.535
Genero        8.61740    1.01354    8.502
Edad        -0.56136    0.15333   -3.661
matricula    0.75792    0.66189    1.145
Estrato      1.20621    0.50064    2.409
trabajo      0.11337    0.02687    4.219
Educacionmadre 0.06794    0.11628    0.584
Educacionpadre 0.19106    0.09942    1.922
Matematicas  1.01565    0.04161   24.410
Metodologia  18.42021    2.94256    6.260
Oficial     -8.11060    2.82657   -2.869

```

Fuente: Los autores - resultado del modelo en lenguaje de programación R.

Ilustración 13 - Ejemplo de Modelo hecho en Python.

```

Mixed Linear Model Regression Results
=====
Model:                MixedLM  Dependent Variable: Razona_cuantitat
No. Observations:    2897      Method:                REML
No. Groups:          63        Scale:                 479.7720
Min. group size:    10        Likelihood:           -13100.7217
Max. group size:    222      Converged:             Yes
Mean group size:    46.0

-----+-----
              Coef.  Std.Err.  z      P>|z|  [0.025  0.975]
-----+-----
Intercept    96.887    5.527  17.529  0.000   86.054  107.721
Genero        8.611    1.015   8.483  0.000    6.622  10.600
Edad       -0.567    0.151  -3.745  0.000   -0.864  -0.270
matricula    0.729    0.675   1.081  0.280   -0.593   2.051
Estrato      1.188    0.507   2.341  0.019    0.193   2.182
trabajo      0.114    0.027   4.227  0.000    0.061   0.166
Educacionmadre 0.068    0.116   0.588  0.557   -0.160   0.297
Educacionpadre 0.192    0.100   1.925  0.054   -0.004   0.387
Matematicas  1.013    0.042  23.946  0.000    0.931   1.096
Metodologia  18.588    3.036   6.122  0.000   12.637  24.538
Oficial     -8.027    2.895  -2.773  0.006  -13.701  -2.353
Group Var    54.744    0.613
=====

```

Fuente Los autores - resultado del modelo en lenguaje Python.

9. CONCLUSIONES

Se automatizó el proceso para el cálculo del valor agregado que las universidades aportan a sus estudiantes, utilizando en el software el método estadístico adoptado por la Universidad Católica de Colombia, reduciendo así el tiempo de trabajo estimado de 240 horas a un máximo de 5 minutos entre la parte computacional y la escogencia del mejor modelo estadístico por parte del investigador.

Para lo anterior se utilizó la metodología RAD que permitió el desarrollo en el tiempo estimado, cumpliendo con la totalidad de los requerimientos funcionales declarados en el documento IEEE 830 de 1988. En base a los requerimientos se realizó el diseño gráfico y funcional del software.

Se programó cada mockup del diseño con el framework Web2Py. Esta interfaz gráfica se integró con los scripts que conforman el BackEnd desarrollados según el diagrama de componentes y el diagrama de ejecución, completando así el desarrollo del software y mostrando al usuario una interfaz gráfica amigable y fácil de manipular.

Se alojó todo el desarrollo en un servidor de AWS; plataforma que brinda seguridad, fiabilidad, escalabilidad y alta disponibilidad; asegurando el acceso desde cualquier dispositivo que cuente con una conexión a internet y la integridad de la información que se encuentra en el servidor.

Realizando las pruebas de caja blanca se comprueba el correcto funcionamiento de cada uno de los componentes del software y con las pruebas de caja negra, se garantiza que los resultados entregados por el software son consistentes con los resultados obtenidos de forma manual por funcionarios de la Universidad Católica de Colombia para la competencia de razonamiento cuantitativo y el grupo de referencia de ingeniería de sistemas y computación, utilizando los resultados de la prueba Saber Pro 2016.

10. RECOMENDACIONES

El funcionamiento del software, en el cargue de las bases, trabaja con una única estructura la cual otorga el ICFES, estructura que en el momento que el mismo organismo las cambie, puede generar errores. Se recomienda la configuración de manera dinámica para el cargue de las bases.

11. TRABAJOS FUTUROS

- Desarrollo de un proceso ETL que descargue automáticamente las bases que cada año el ICFES sube a su FTP y sean cargadas a la base de datos en la estructura correspondiente para el funcionamiento del software.
- Desarrollar un nuevo módulo que calcule el valor agregado en competencias específicas.
- Diseñar y desarrollar un módulo que calcule el valor agregado por otras metodologías.

12. BIBLIOGRAFÍA

Modelo de Valor Agregado: una implementación para el caso de la educación superior en Colombia, Isabella Muñoz Gómez, Universidad Nacional de Colombia, Facultad de Ciencias Económicas, Maestría en Ciencias Económicas, 2016

Valor agregado como indicador de calidad en programas académicos universitarios utilizando resultados de pruebas censales, Marco Tulio Rodríguez Sandoval & Rafael Alberto Vilchez Pirela, Corporación Universitaria del Caribe CECAR, Colombia, Revista Academia & Virtualidad 10(1): 42-55, 2017

Medición del valor agregado para la educación superior en Bogotá, Ramiro Rodríguez Revilla, Universidad Santo Tomás, Facultad de Ciencias Económicas y Administrativas, Maestría en Ciencias Económicas, 2015

La intervención académica en la construcción de una sociedad con calidad: análisis del valor agregado en el proceso formativo colombiano, Carlos-Hernán Isáziga-David, Jesús Gabalán-Coello, Fredy-Eduardo Vásquez-Rizo, HALLAZGOS / ISSN: 1794-3841 / Año 11, N.º 22 / Bogotá, D. C. / Universidad Santo Tomás / pp. 359-384, 2014

Análisis de los resultados de las Pruebas SABER 11 y SABER PRO, Julio Cesar Alonso, Paola Casasbuenas, Beatriz Gallo, Giselle Torres, Cienfi - Universidad Icesi, 2012

BARRERO CALIXTO, Juan Carlos, Modelo de Valor Agregado para la carrera de Ingeniería de Sistemas y Computación a partir de Saber PRO 2015 y 2016. Pereira, 2018. Universidad Tecnológica de Pereira UTP, Facultad de Ciencias Empresariales

BUKOUSKI, Pawel; KOBUS, Martina; Economics of Education Review: The threat of competition and public school performance: Evidence from Poland. December 2018. London School of Economics and Political Science, ScienceDirect URL: <https://www.sciencedirect.com/science/article/pii/S0272775717306878>

GALLARDO ERASO, Luis Álvaro; RINCÓN QUIÑONES, Carlos Gustavo, Pedagogía y Saberes No. 35: La desigualdad universitaria, una realidad nacida del mito de la eficiencia. Bogotá, 2011, 14 h. Universidad Pedagógica Nacional. Facultad de Educación URL: <http://revistas.pedagogica.edu.co/index.php/PYS/article/download/948/963/>

ESTEBAN, Belkys P. Educación de calidad: Se renueva la institución del Icfes. 2010. Ministerio de Educación Nacional de Colombia. URL: <https://www.mineducacion.gov.co/observatorio/1722/article-234655.html>

ICFES, mejor saber – Valor agregado y Aporte relativo – Aporte Relativo. 2019. URL: <http://www.icfes.gov.co/aporte-relativo>

Ministerio de Educación – Evaluación de Competencias Ciudadanas. 2012. Margarita Peña Borrero- ICFES. URL: https://www.mineducacion.gov.co/cvn/1665/articles-314094_archivo_pdf_6.pdf

Universidad ICESI – Resultados de Icesi por GRUPOS DE REFERENCIA. URL: <http://www.icesi.edu.co/saberpro2018/index.php/resultados/por-grupos-de-referencia>

PROCEDIA - Social and Behavioral Sciences Volume 176, 20 February 2015, Pages 940-945 - Formative Research in Higher Education: Some Reflections. URL: <https://www.sciencedirect.com/science/article/pii/S1877042815005996>

Pontificia Universidad Javeriana Cali, ¿Qué es Saber Pro? URL: <https://www.javerianacali.edu.co/admisiones/saberpro/que-es-saber-pro>

PERKTOLD, J. O. S. E. F., Seabold, S. K. I. P. P. E. R., & Taylor, J. O. N. A. T. H. A. N. (s.f.). StatsModels: Statistics in Python — statsmodels 0.9.0 documentation. Recuperado 20 abril, 2019, de <https://www.statsmodels.org/stable/index.html>

DI PIERRO, M. A. S. S. I. M. O., Mulone, M. A. R. T. I. N., & Maldonado, J. E. N. N. I. F. E. R. (s.f.). web2py - Introducción. Recuperado 21 abril, 2019, de <http://www.web2py.com/books/default/chapter/36/01/introduccion>

STEPHEN W. Raudenbush, Michigan State University, East Lansing – Hierarchical Linear Models and Experimental Design. URL: http://jakewestfall.org/misc/Raudenbush_1993.pdf

ZUUR, A., Ieno, E., & Smith, G. (2007). *Analysing Ecological Data* (p. 125). Dordrecht: Springer.

CORREA Morales, J., & Salazar Uribe, J. (2016). *Introducción a los modelos mixtos* (1st ed., pp. 11-241). Medellín, Colombia.

PINHEIRO, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2013). nlme: Linear and nonlinear mixed effects models. R package version, 3(1), 111.

MÁRQUEZ CEBRIÁN, M. D. (2002). *Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificación*. Universitat Politècnica de Catalunya.

SCHWARZ, Gideon E. (1978). "Estimating the dimension of a model". *Annals of Statistics* 6 (2): 461–464. doi:10.1214/aos/1176344136. MR468014.

1 Calderón Rivera, D., Navarrete López, C., & Díaz Arévalo, J. (2016). Ajustes de distribuciones probabilísticas para la variable temperatura media multianual para el departamento de Boyacá (Colombia). *Ingeniería Y Región*, 14(2), 113-124.

Universidad Autónoma del Caribe, Revisión de metodologías ágiles para el desarrollo de software. *PROSPECTIVA*, vol. 11, núm. 2, julio-diciembre, 2013, pág. 30-39. URL: <https://www.redalyc.org/pdf/4962/496250736004.pdf>

JAMES Martin, (1991) *Rapid Application Development*; Macmillan Publishing Co., Inc.