

PROPUESTA DE MODELO PREDICTIVO APLICADO AL RENDIMIENTO EN
NADADORES COLOMBIANOS

PAULA ANDREA FIGUEROA POLANCO.

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
EN CONVENIO CON LA UNIVERSIDAD ABIERTA DE CATALUÑA
FACULTAD DE INGENIERÍA
SANTIAGO DE CALI
2017

PROPUESTA DE MODELO PREDICTIVO APLICADO AL RENDIMIENTO EN
NADADORES COLOMBIANOS

PAULA ANDREA FIGUEROA POLANCO

Trabajo de grado para optar por el título de Magister en Software Libre

Ph.D. Carlos Mauricio Gaona Cuevas
Profesor Titular Universidad del Valle
Director de tesis

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
EN CONVENIO CON LA UNIVERSIDAD ABIERTA DE CATALUÑA
FACULTAD DE INGENIERÍA
SANTIAGO DE CALI
2017

Nota de aceptación

Presidente del jurado

Jurado

Jurado

Bucaramanga, 20 de octubre de 2017

Agradecimientos

La autora expresa su más sincero agradecimiento a:

A mi madre a quien le debo todo lo que soy.

El profesor Carlos Mauricio Gaona, director del proyecto de tesis, cuyos aportes y disposición constantes fueron fundamentales para concluir en buen término la presente investigación.

A la profesora Claudia por sus orientaciones en la asignatura de Tesis y la profesora Liliana por sus valiosos aportes que me ayudaron a complementar mi formación y esta investigación.

CONTENIDO

	pág.
RESUMEN	1
INTRODUCCIÓN	2
1. ASPECTOS GENERALES	4
1.1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN	4
1.1.1. FORMULACIÓN DEL PROBLEMA CENTRAL	4
1.1.2. FORMULACIÓN DE PROBLEMAS COMPLEMENTARIOS	4
1.1.3. PLANTEAMIENTO DEL PROBLEMA	4
1.1.4. JUSTIFICACIÓN	6
1.2. DELIMITACIÓN	6
1.2.1. TEMPORAL Y GEOGRÁFICA	6
1.2.2. CONCEPTUAL	7
1.3. OBJETIVOS	7
1.3.1. OBJETIVO GENERAL	7
1.3.2. OBJETIVOS ESPECÍFICOS	8
2. MARCO TEÓRICO	9
2.1. ANTECEDENTES	9
2.2. MEDICIÓN DEL RENDIMIENTO EN LA NATACIÓN	10
2.2.1. LA VELOCIDAD CRÍTICA COMO INDICADOR DE RENDIMIENTO	10
2.2.2. EL FACTOR FATIGA EN LA ECUACIÓN DE RESISTENCIA	12

2.2.3.	RITMOS ESTIMADOS POR NIVEL	13
2.3.	DATOS, MODELOS PREDICTIVOS, DISEÑO Y TÉCNICAS	14
2.3.1.	ANÁLISIS Y TRATAMIENTO DE DATOS	14
2.3.2.	MODELOS PREDICTIVOS	15
2.3.3.	MODELOS DE REGRESIÓN LINEAL SIMPLE	16
2.4.	SOFTWARE Y MODELOS DE PREDICCIÓN EXISTENTES APLICADOS A DEPORTES	17
2.4.1.	APLICACIÓN DE UN MODELO MATEMÁTICO AL TAPER EN JÓVENES NADADORES	17
2.4.2.	NUMERICAL ALGORITHMS FOR PREDICTING SPORTS RESULTS	18
2.4.3.	A HYBRID PREDICTION SYSTEM FOR AMERICAN NFL RESULTS	19
2.4.4.	SISTEMA DE PREDICCIÓN DE RESULTADOS EN EVENTOS DEPORTIVOS Y SU APLICACIÓN EN LAS APUESTAS	19
3.	ESTRATEGIA METODOLÓGICA	21
3.1.	MÉTODO DE INVESTIGACIÓN	21
3.2.	DESCRIPCIÓN DE LA METODOLOGÍA	21
3.2.1.	PRIMERA FASE: GENERACIÓN DE LA PROPUESTA	22
3.2.2.	SEGUNDA FASE: RECOPIACIÓN, ANÁLISIS Y COMPRESIÓN DE LOS DATOS	23
3.2.3.	TERCERA FASE: DISEÑO DEL MODELO	24
3.2.4.	CUARTA FASE: IMPLEMENTACIÓN Y DESPLIEGUE	24
4.	RESULTADOS DE LA INVESTIGACIÓN	26
4.1.	PRIMERA FASE: GENERACIÓN DE LA PROPUESTA	26

4.2.	SEGUNDA FASE: RECOPIACIÓN, DESCRIPCIÓN, EXPLORACIÓN Y PREPARACIÓN DE LOS DATOS	26
4.2.1.	RECOPIACIÓN Y DESCRIPCIÓN DE LOS DATOS	26
4.2.2.	EXPLORACIÓN DE LOS DATOS	27
4.3.	TERCERA FASE: SELECCIÓN DE LA TÉCNICA DE MODELADO	29
4.3.1.	ANÁLISIS DE DATOS CON R.	29
4.3.2.	DISEÑO DEL MODELO	34
4.3.3.	EVALUACIÓN DEL MODELO	36
4.4.	CUARTA FASE: IMPLEMENTACIÓN DEL MODELO	40
4.4.1.	DEFINICIÓN DE REQUERIMIENTOS	40
4.4.2.	DISEÑO DE PROTOTIPO	43
4.4.3.	IMPLEMENTACIÓN	45
4.4.3.1.	SELECCIÓN DE TECNOLOGÍAS:	45
4.4.3.2.	DISEÑO DE BASE DE DATOS	46
4.4.3.3.	DISEÑO DE PLANTILLAS O TEMPLATES	47
4.4.3.4.	DISEÑO DE VISTAS	51
4.4.3.5.	PRUEBAS FUNCIONALES	55
5.	CONCLUSIONES	59
6.	RECOMENDACIONES Y TRABAJOS FUTUROS	61
	REFERENCIAS Y BIBLIOGRAFÍA	62
	ANEXOS	67

LISTA DE TABLAS

	pág.
Tabla 1 - Datos recolectados para hombres	28
Tabla 2 - Datos recolectados para mujeres	28
Tabla 3 - Cálculo de la precisión del modelo a través del MAPE con muestras aleatorias	39
Tabla 4 - Cálculo de la precisión del modelo a través del MAPE con muestras con distribución uniforme en la clasificación de las distancias	39
Tabla 5 - Lista de requerimientos	41
Tabla 6 - Comparativo de los resultados registrados por los nadadores y los obtenidos en la predicción	58

LISTA DE ECUACIONES

	pág.
Ecuación 1: Cálculo de velocidad crítica	11
Ecuación 2: Cálculo de la distancia a partir de CSS y ASC	11
Ecuación 3: Ecuación de resistencia	12
Ecuación 4: Regresión Lineal	16
Ecuación 5: Cálculo de la pendiente	16
Ecuación 6: Cálculo del intercepto	16
Ecuación 7: Coeficiente de determinación	17
Ecuación 8: Coeficiente de correlación	17
Ecuación 9: Cálculo del tiempo a partir del CSS y ASC	35
Ecuación 10: Fórmula del cálculo del tiempo planteada por Riegel	35
Ecuación 11: Cálculo de la pendiente aplicando transformación de los datos	35
Ecuación 12: Transformación del cálculo del intercepto	36
Ecuación 13: Suma acumulada de errores de pronóstico	37
Ecuación 14: Desviación media absoluta MAD	37
Ecuación 15: Error cuadrático medio MSE	38
Ecuación 16: Error porcentual medio absoluto MAPE	38

LISTA DE FIGURAS

	pág.
Figura 1. Ficha bibliográfica	23
Figura 2. Información almacenada en FECNA	27
Figura 3. Carga de todos los datos y variables	30
Figura 4. Resumen estadístico arrojado por R	30
Figura 5. Análisis de correlación de los datos	31
Figura 6. Estudio de la distribución normal a partir de Lilliefors.	31
Figura 7. Comportamiento del tiempo (resultado) vs la prueba (distancia).	32
Figura 8. Gráficas de normalidad del tiempo en distancias 100, 200, 400, 800 y 1500 mts.	33
Figura 9. Mejor modelo de regresión y gráfica adaptada al comportamiento de los datos.	34
Figura 10. Modelo implementado en R.	36
Figura 11. Plantilla para registro de HU.	40
Figura 12. Interfaz para listar pruebas	43
Figura 13. Formulario para registro y edición de prueba	43
Figura 14. Interfaz para confirmación de eliminación de registro	44
Figura 15. Interfaz para carga masiva	44
Figura 16. Interfaz para visualización de resultados	45
Figura 17. Modelo Template Vista de Django	46

Figura 18. Modelo de base de datos.	47
Figura 19. Formulario de ingreso de pruebas	48
Figura 20. Carga masiva de registros	49
Figura 21. Confirmación de eliminación de un registro	49
Figura 22. Listado de pruebas	50
Figura 23. Informe de resultados	50
Figura 24. Instructivo de uso del aplicativo	51
Figura 25. Funciones listar, crear, editar y borrar pruebas	52
Figura 26. Captura de datos a través de las vistas	52
Figura 27. Ejemplo de cálculo de la media a partir de la distancia	52
Figura 28. Cálculo de la pendiente y el intercepto	53
Figura 29. Tiempo estimado, ritmo estimado, coeficiente de determinación	53
Figura 30. Cálculos aplicando a la ecuación de resistencia	53
Figura 31. Ritmo y tiempo estimado haciendo uso del factor fatiga	54
Figura 32. Cálculo de ritmos por nivel	54
Figura 33. Definición para convertir a horas, minutos y segundos	54
Figura 34. Marcas de nadadores 1, 2 y 3	55
Figura 35. Resultados según proyección del nadador 1 para 1500mts	56
Figura 36. Resultados según proyección de nadador 2 para 800 mts	56
Figura 37. Resultados según proyección de nadador 3 para 400mts	57
Figura 38. Datos tomados de la FECNA Nadadores 1, 2 y 3	57

LISTA DE ANEXOS

	pág.
Anexo 1 Formato de historias de usuario	67

RESUMEN

La presente investigación se enfocó en el diseño de un modelo predictivo para nadadores a partir de los datos registrados en la Federación Colombiana de Natación (FECNA). La implementación del modelo predictivo se realizó mediante una aplicación web bajo las políticas del software libre y licencia de *Creative Commons*.

La estimación o predicción del tiempo del atleta se desarrolló a través de dos modelos: en el primero se implementa el modelo de regresión lineal a la ecuación de velocidad crítica (*Critical Swim Speed - CSS*) propuesta por Wakayoshi en el que se estima el tiempo a partir del cálculo de los valores de CSS y la capacidad de natación anaeróbica (*Anaerobic Swimming Capacity - ASC*). En el segundo modelo se usa la ecuación de resistencia planteada por Riegel para realizar una estimación del tiempo a partir del factor fatiga partiendo del cálculo de dos constantes básicas teniendo en cuenta la relación entre el tiempo y la distancia: $\text{Tiempo} = b * \text{distancia}^m$. Para medir la precisión del modelo se hizo uso del error porcentual medio absoluto donde se obtiene que los modelos son bastante acertados en sus estimaciones pues los porcentajes de error en ambos están por debajo del 20%.

La aplicación web da un reporte de valores estimados que puede ser usados por el atleta para analizar su rendimiento a partir de una distancia objetivo; se realizó con el framework *Django* y el lenguaje de programación *Python* y para que la aplicación tuviera interfaces adaptables a cualquier dispositivo se usó el Framework *Bootstrap*.

Por último las pruebas funcionales en la plataforma consistieron en plantear tres escenarios de prueba tomando los datos históricos de la FECNA, se pudo obtener valores donde la estimación del modelo se acercaba en un alto porcentaje a las marcas obtenidas por el atleta.

Palabras claves: modelo predictivo, natación, aplicación web, fecna, tiempo estimado, distancia, pruebas, marcas.

INTRODUCCIÓN

El crecimiento constante de las tecnologías de la información y las comunicaciones (TIC) y su aplicación interdisciplinar le ha permitido involucrarse en distintos aspectos de la vida, es una herramienta que combinada con otras especialidades, con el incremento constante de la información y con su almacenamiento, son de gran utilidad a la hora de obtener y procesar información de fenómenos que ocurren a diario.

Los deportes han tomado gran protagonismo en la era actual, los atletas son preparados desde pequeños y su rendimiento es aprovechado a través de entrenamientos adaptados a sus necesidades o a las capacidades que van desarrollando de acuerdo a su preparación y a las marcas obtenidas en competencias. De esta manera se puede estimar los resultados que obtendrá, por ejemplo un nadador, en determinada prueba. Una estrategia que facilitaría este proceso sería la implementación de un modelo predictivo, el cual se puede ver como el resultado de la combinación de datos y fórmulas matemáticas, donde el aprendizaje que requiere el modelo puede ser traducido en la creación de una función de correlación entre un conjunto de campos de datos de entrada y una respuesta o destino variable (Guazzelli, 2012).

Algunas de las técnicas de modelado predictivo existen para ayudar a traducir estos datos en conocimiento y valor. Para Guazzelli (2012) esto se logra por medio de patrones de aprendizaje escondidos en grandes volúmenes de datos históricos. Una vez que se completa el aprendizaje, el resultado es un modelo predictivo. Las técnicas de modelado permiten el desarrollo de modelos predictivos precisos, siempre y cuando existan datos suficientes y la calidad de los datos no sea un problema.

Colombia es un país con niveles muy bajos de cultura de registro de información deportiva y con gran variedad de deportistas en diferentes disciplinas, de lo que cabe destacar la mayoría son empíricos. Es difícil conseguir entrenadores que usen el conocimiento obtenido a partir de los resultados de las pruebas y competencias en las que participan los atletas, como herramienta para mejorar sus entrenamientos y de esta manera explotar el potencial de cada uno de ellos. Esta problemática existe principalmente por que no se tiene información.

De acuerdo a lo anterior, se desarrolló una aplicación web que arroja valores estimados del rendimiento ingresando una velocidad objetivo y haciendo uso de las

ecuaciones de velocidad crítica y resistencia. Estas fueron entrenadas a partir de los resultados registrados de tiempo y distancia de nadadores colombianos en eventos deportivos de natación.

Para desarrollar este sistema se utilizó en su fase inicial los datos registrados en la base de datos de la Federación Nacional de Natación (FECNA), posteriormente se realizó un procesado de los mismos que facilitó el análisis y respectiva aplicación a través de técnicas de regresión lineal, potencial y transformación de datos, a las dos ecuaciones existentes más usadas para medir el rendimiento, estas son la velocidad crítica (CSS) y la ecuación de resistencia (FF). El resultado obtenido es una estimación del tiempo a partir de una distancia objetivo.

Es así como se da inicio a la revisión del estado del arte en cuanto a los modelos predictivos y sus aplicaciones, se describe el proceso para la recolección, transformación y análisis de los datos para pasar a la implementación del prototipo. En esta última se aplican los principios de la metodología de desarrollo ágil *Extreme Programming*, en las que se presenta las principales características de diseño y funcionamiento haciendo uso para el desarrollo de tecnologías de uso libre como *Django*, *Bootstrap*, *Postgres*. Finalmente se realiza una prueba de la plataforma web haciendo uso de tres nadadores, registrados en el histórico de los datos.

1. ASPECTOS GENERALES

1.1. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN

1.1.1. Formulación del problema central

¿Cómo proponer una plataforma web que implemente un modelo predictivo para estimar el rendimiento, visto como el tiempo que demora un nadador en recorrer una distancia?

1.1.2. Formulación de problemas complementarios

- ¿Cómo a través de la información histórica registrada de nadadores colombianos se pueden obtener indicadores que permitan estimar su rendimiento?
- ¿De qué manera a través de ciertas variables aplicadas se puede generar un modelo que realice estimaciones del rendimiento de los nadadores?
- ¿Cuáles son las variables que influyen en el rendimiento de un nadador?
- ¿De qué manera se puede impulsar la utilidad de una herramienta que implemente modelos predictivos en deportes para entrenadores y deportistas en el área de natación?

1.1.3. Planteamiento del problema

La presente investigación se enfocó en el diseño de un modelo predictivo para estimar el rendimiento deportistas en el área de natación a partir de los datos registrados de distancia y tiempo en la Federación Colombiana de Natación y de las técnicas conocidas para medir el rendimiento de los nadadores como son la velocidad crítica y la ecuación de resistencia.

El origen de este planteamiento radica en la necesidad existente hoy en día de ofrecer un modelo que permita mejorar los entrenamientos de los nadadores colombianos teniendo en cuenta que cada día la participación de los mismos en eventos deportivos nacionales e internacionales es más destacada y común, además que existen muchos datos históricos que se van incrementando con el tiempo, y que ofrecen información valiosa y de gran relevancia para plantear estrategias de mejoramiento en sus entrenamientos.

Los modelos predictivos se desarrollan a través de patrones de aprendizaje que surgen del análisis de grandes volúmenes de datos históricos, posteriormente se valida el modelo y de esta manera se garantiza que sea capaz de generalizar el conocimiento que ha aprendido y aplicarlo a una nueva situación (Guazzelli, 2012). La aplicación de múltiples estrategias desarrolladas en distintas áreas, matemáticas o estadísticas, para el análisis de datos permite explotar el potencial de los datos y al sistematizarlo ofrece disponibilidad a quienes puedan aprovecharlo.

En el mundo existen muchos sistemas para realizar pronósticos que se valen de distintas herramientas matemáticas o estadísticas para tratar los datos y obtener estimados sobre los resultados obtenidos en un partido, estos comúnmente se centran en deportes como fútbol, fútbol americano, baloncesto, tenis entre otros y se relacionan más con el hecho de ofrecer un software para realizar apuestas seguras.

Actualmente en Colombia, se ha iniciado un proceso de integración y registro de información de resultados de competencias deportivas de diferentes deportes tales como natación, atletismo, etc., sin embargo a la fecha estos resultados deportivos no son usados por los entrenadores dado que no se cuenta con herramientas que arrojen cierto tipos de pronósticos para eventos deportivos, especialmente en la natación ya que los entrenadores se basan en su experiencia o en los conocimientos previos que tienen para aprovechar en mejor medida las capacidades de sus deportistas.

Se espera que el modelo planteado sirva como herramienta para planificar entrenamientos y explotar las capacidades de los nadadores de alto rendimiento en Colombia.

1.1.4. Justificación

En la actualidad existen muchos sitios dedicados a predicciones de las competencias deportivas y deportes más populares tales como fútbol, fútbol americano, baloncesto, hockey y tenis. Estos sitios plantean algoritmos basados en datos históricos de cada equipo y proponen posibles ganadores partiendo del sitio donde se realizará el evento, si el equipo es local o visitante y el porcentaje de partidos ganados con el equipo al que se enfrenta. Estos sitios son desarrollados para que sean usados por fanáticos y apostadores y es debido a este enfoque que no ofrecen información que sea de utilidad para entrenadores y atletas que les permita evaluar a partir de ciertas condiciones como deben enfocar sus entrenamientos para aprovechar sus habilidades y fortalezas.

Por otro lado en nuestro país a pesar de la gran participación existente de múltiples deportistas en diferentes eventos de manera sobresaliente no se ha implementado una estrategia de análisis de entrenamiento formal basada en los datos más que en la experiencia. El análisis de datos históricos permita formar futuros deportistas y grupos de ellos fortalecidos a partir de la información inicial y conociendo con anterioridad en que momento su habilidad es máxima o cuánto tiempo/entrenamiento se requiere para alcanzar su punto ideal antes de volver a descender.

Frente a esto se hace necesario ofrecer a la comunidad especialista en esta área un mecanismo con la capacidad suficiente para administrar y recolectar toda la información referente a los atletas, procesarla y mostrar posibles resultados a partir de ella. Este sistema deberá proporcionar la opción para ingresar los datos y estar soportado por diferentes estrategias estadísticas para tratarlos mostrando al usuario información gráfica y estadística que le permita valorar el estado actual de su entrenamiento además de la posibilidad de estar disponible en el SIND.

1.2. DELIMITACIÓN

1.2.1. Temporal y geográfica

El presente proyecto tuvo una duración de 8 meses en la ciudad de Cali del departamento del Valle del Cauca.

El proyecto incluye dos fases: la primera relacionada con la revisión y recolección de información disponible y la segunda la obtención de un estado del arte que sirviera como herramienta para fase de diseño, análisis e implementación del modelo predictivo en la aplicación web.

1.2.2. Conceptual

El presente proyecto presenta una aplicación web en la que se implementa el modelo predictivo y sirve como herramienta para que nadadores y entrenadores puedan hacer un análisis del rendimiento a través de la medida de diferentes aspectos que interfieren en el desempeño deportivo de un nadador como son la velocidad crítica, el factor fatiga, los ritmos por nivel de Phil Skiba y Bill Sweetenham.

Este proyecto vincula los conceptos relacionados con natación con un aporte del desarrollo de software y aplicaciones web para ofrecer una herramienta de acceso web de fácil uso, con interfaces adaptables y adaptable donde el nadador pueda obtener un resumen claro que le permita analizar los indicadores de rendimiento ingresando sus marcas y la distancia que desea recorrer.

Este modelo se entrenó con una muestra de los datos registrados en la FECNA y se generó aplicando técnicas estadísticas en ella con el fin de conocer su comportamiento y determinar una curva que mejor se ajustará para estimar los resultados de acuerdo a la muestra. Al tener generado el modelo se pasó a realizar la implementación del software y se probó con otros datos reales que permitieran determinar la exactitud de la predicción.

1.3. OBJETIVOS

1.3.1. Objetivo general

Desarrollar una aplicación web que implemente un modelo de predicción aplicado al rendimiento de nadadores colombianos.

1.3.2. Objetivos específicos

1. Realizar la recolección y caracterización de los datos a partir de la información almacenada en la Federación Colombiana de Natación.
2. Definir las variables o indicadores que influyen en el rendimiento en natación.
3. Diseñar un modelo aplicado al rendimiento de nadadores colombianos para realizar pronósticos en competencias deportivas.
4. Desarrollar una plataforma web de software libre que permita validar el modelo con la posibilidad de integrarlo al SIND.

2. MARCO TEÓRICO

2.1. ANTECEDENTES

Para iniciar con el diseño e implementación del modelo de predicción para nadadores colombianos se partió de varios conceptos como: velocidad crítica planteada por Wakayoshi, curva de fatiga planteada por Riege, ritmos, ritmos por nivel según Phil Skiba o Bill Sweetenham y regresión lineal (RL). La claridad en esta conceptualización se hace necesaria para el entendimiento de los factores que influyen directamente en el desempeño y tiempos de ejecución de pruebas en los nadadores.

En la implementación de este tipo de modelos se hace necesario la obtención de datos históricos que deben ser procesados a través de diferentes técnicas estadísticas, tales como cálculo de la media, estudio de la varianza o la desviación estándar y así observar el comportamiento de las variables que definirán los pronóstico respecto al tiempo del atleta en determinada prueba.

En la literatura se pueden encontrar estudios similares aplicados a varias disciplinas del conocimiento que pretenden realizar predicciones principalmente en deportes como fútbol, baloncesto o fútbol americano. Estas predicciones están enfocadas esencialmente en obtener resultados deportivos basados en apuestas y aplicación de técnicas matemáticas y estadísticas como regresión lineal, redes neuronales, teorema de bayes entre otros. Aunque existen modelos aplicados a la natación no están implementados en herramientas que faciliten a entrenadores y deportistas generar estrategias para sacar mejor provecho de sus entrenamientos.

Además de presentarse modelos en investigaciones también existen páginas y blogs especializados en el tema de predicción deportiva pero que no presentan ninguna base matemática bajo la cual muestran la información, aunque se puede evidenciar que pueden ser personas expertas de manera empírica en el tema o aficionados los cuales proponen apuestas que según su experiencia son más seguras.

Con los trabajos y software más destacados que se describen en esta sección se intenta obtener una idea de cómo realizar los procesos de recolección y que estrategias de predicción son útiles para cumplir los objetivos, además de tener una idea de las aplicaciones actuales en el mercado.

La presente sección se encuentra dividida en tres partes la primera relacionada con la base teórica de conceptos directamente relacionados con la medición del rendimiento en la natación, en la segunda encontrará conceptualización sobre modelos de regresión lineales y en la tercera la descripción de aplicaciones, software y páginas disponibles relacionados con modelos predictivos y las estrategias realizadas para la obtención de los mismos.

2.2. MEDICIÓN DEL RENDIMIENTO EN LA NATACIÓN

2.2.1. La velocidad crítica como indicador de rendimiento

Las competencias de natación básicamente tienen como objetivo cubrir una distancia en el menor tiempo posible aplicando cierto tipo de técnicas en las que el nadador se ha hecho experto. Se debe aclarar que las técnicas de nado son conceptos muy diferentes a los de estilo, puesto que la técnica son movimientos o patrones que buscan aprovechar el potencial del deportista. Entre las técnicas más conocidas se tiene: libre, espalda, pecho y mariposa. Mientras que el estilo es la impresión personal que le da cada nadador al ejecutar la técnica, es decir dos nadadores pueden ser expertos en la misma técnica pero aplicar un estilo diferente que incluye el ritmo o movimiento que imprime a su técnica.

El tiempo de la prueba tradicionalmente se mide en un cálculo entre el tiempo promedio y el tiempo final junto con ciertas observaciones técnicas que más que nada son cualitativas, al añadir otras variables como la edad, el sexo, la longitud del ciclo, el ritmo, la velocidad crítica o el factor de fatiga, permiten obtener información más detallada del rendimiento del atleta y realizar un estimado de cómo será su comportamiento en determinada competencia.

Para que los entrenadores puedan tener claridad sobre el rendimiento de los deportistas se hace necesario realizar una planificación en los entrenamientos, esta actividad y especialmente en natación es una de las tareas más complejas, puesto que esta varía de las condiciones de un atleta a otro (Wakayoshi K., 2012), es por esta razón que expertos en el área propone la velocidad crítica como un método muy eficaz para la programación de entrenamientos.

“La velocidad crítica se define como la velocidad máxima de nado que se puede mantener durante un largo período de tiempo sin que el nadador llegue a la

extenuación. Se expresa como la pendiente de una recta entre la distancia de nado y el tiempo en cubrir esa distancia” (Zingsem, 2015).

Ecuación 1: Cálculo de velocidad crítica

$$Velocidad\ critica = \frac{d_2 - d_1}{t_2 - t_1}$$

Fuente:

<http://g-se.com/es/entrenamiento-en-natacion/blog/uso-de-la-velocidad-critica-para-el-entrenamiento-de-la-resistencia-aerobica-en-nadadores-jovenes>

El uso de más de tres marcas reduce el error en el cálculo de la velocidad crítica y la toma de las mismas implican que se debe realizar un máximo esfuerzo y deben existir tiempos largos de descansos entre cada prueba.

Para Wakayoshi et al. (2012) se puede plantear la velocidad crítica como un modelo de predicción de rendimiento realizando varias tomas experimentales y convirtiéndolos en una relación lineal donde la pendiente obtenida representaba la CSS y la ordenada al origen la ASC.

Ecuación 2: Cálculo de la distancia a partir de CSS y ASC

$$Distancia = CSS * Tiempo + ASC$$

Fuente:

<http://www.amtriathlon.com/2007/01/prediccin-del-rendimiento-en-natacin.html>

Esta prueba es altamente usada en nadadores porque no es invasiva, es de bajo costo y permite obtener medidas de la capacidad aeróbica y anaeróbica (Martinez A. , 2007).

Adicionalmente la literatura sugiere que la velocidad crítica puede ser utilizada para determinar los tiempos del entrenamiento aplicados a series en diferentes distancias a través de la especificación de las intensidades del tratamiento basado en los cálculos de la CSS, cada uno de estos valores cambia dependiendo de las condiciones del nadador (Coulson & CooperD., 2011).

2.2.2. El factor fatiga en la ecuación de resistencia

Uno de los aspectos necesarios para la estimación del rendimiento es medir la resistencia, la cual indica la capacidad que tiene el atleta a resistir fatiga, esta relación es inversamente proporcional, entre más alta sea la fatiga menor será resistencia.

La ecuación de resistencia requiere solo el conocimiento de dos constantes que permiten estimar el comportamiento de la distancia, el tiempo y la velocidad. El objetivo de la ecuación es estimar el ritmo al que puede llegar un nadador partiendo de la relación que existe entre diferentes tiempos para las diferentes pruebas realizadas.

Para el cálculo del factor de fatiga se tienen en cuenta dos aspectos: “La curva de economía de carrera que es la velocidad de carrera con el consumo de oxígeno necesario para sostenerla y la curva de duración-intensidad: relaciona la duración de la prueba con la proporción del consumo máximo de oxígeno que se puede mantener” (Martínez A. E., 2012). Para mejorar una mejor estimación de la ecuación se hizo uso de mínimos cuadrados a los logaritmos de tiempo vs distancia.

Actualmente el factor fatiga se maneja como una ecuación exponencial de la forma que se expresa en la ecuación 3:

Ecuación 3: Ecuación de resistencia

$$Tiempo = m + distancia^b$$

Fuente:

http://www.amtriathlon.com/2007/01/prediccin-del-rendimiento-en-natacin_29.html

Donde m es la velocidad media calculada a partir del estimado de las marcas registradas y b es el factor fatiga que determina el comportamiento con el que disminuye la velocidad media dada la distancia y el tiempo. Entre más grande sea la velocidad y la distancia, menos resistente será el atleta a la fatiga lo que hará aumentar el tiempo (Riegel, 1977). La medida mundial para el factor de fatiga en natación para hombres y mujeres es el mismo con un valor de 1.03.

2.2.3. Ritmos estimados por Nivel

Las mediciones de rendimiento, tiempo y velocidad en la natación tienen una fuerte relación con la alta densidad del agua que afecta la potencia que se requiere para aumentar la velocidad en una relación directamente proporcional, es decir cuando la potencia aumenta la velocidad aumenta y si disminuye también lo hace la velocidad. Este esfuerzo en añadir más potencia implica un desgaste calórico por lo que las modificaciones, así sean pequeñas en la velocidad modificará el nivel de resistencia del nadador, es por esta razón que entrenadores recomiendan mantener una relación constante de velocidad/ritmo en natación para que no afecte la potencia en unidades de tiempo y que creen una diferencia en la marca.

Martínez (2007) refiere que los ritmos por nivel se deben especificar a partir de la función de la velocidad umbral funcional, la cual representa la máxima potencia promedio que puede sostener un atleta durante un tiempo máximo aproximado. Lo que en natación específicamente se conoce como velocidad crítica (CSS).

Los niveles en los ritmos fueron propuestos por Dr. Andrew Coggan, planteados inicialmente para ciclismo, y en natación los entrenadores Phil Skiba y Bill Sweetenham los adaptaron basándose en la fisiología del ejercicio. El sistema consta de 7 niveles los cuales Martínez (2007) describe así:

Nivel 1 – Recuperación Activa: la velocidad es menor 93% de la velocidad crítica, se debe aumentar la circulación para facilitar los procesos de recuperación y reparación, descanso sin inactividad. Para Sweetenham es la zona 1 ejercicios aeróbicos de baja intensidad representados como A1.

Nivel 2 – Resistencia Aeróbica: la velocidad debe estar entre 93-97% de la velocidad crítica. Mejora la resistencia a la fatiga y la duración es más importante que la intensidad. Para Sweetenham se refiere al Mantenimiento aeróbico A2 de la zona 1.

Nivel 3 – Tempo: La velocidad está entre 97-99% de la velocidad crítica. Para el caso de de Swetenham en la zona 1 A3 refiere al entrenamiento aeróbico intensivo o desarrollo aeróbico.

Nivel 4 – Umbral: El rango de velocidad se comprende entre 99-102% el objetivo es maximizar las adaptaciones metabólicas a nivel muscular, que son las que más se afectan al recorrer las distancias en competencias de resistencia/fondo.

Nivel 5 – VO₂max: El rango de velocidad se encuentra entre 102-109% de la velocidad crítica. Se relaciona con la maximización de las adaptaciones cardiovasculares, que afectan el recorrido de distancias de competencias de resistencia/fondo.

Nivel 6 – Capacidad Anaeróbica: La velocidad es mayor a 109%, debe mejorar en distancias cortas pues directamente la resistencia anaeróbica.

Nivel 7 – Sprint: no tiene relación con la velocidad crítica.

2.3. DATOS, MODELOS PREDICTIVOS, DISEÑO Y TÉCNICAS

2.3.1. Análisis y tratamiento de datos

Cuando se realizan procedimientos que requieren el manejo y análisis de datos, como es el caso de los modelos predictivos, se hace necesario implementar pruebas que permitan determinar el comportamiento de los mismos para así garantizar la fiabilidad de estos modelos.

Para dicho estudio existen las pruebas de normalidad, las cuales permiten determinar si un conjunto de datos se ajusta a una probabilidad determinada, es decir que si tiene una simetría alrededor de la media y si esta coincide con la mediana. Se aplica la distribución normal, especialmente en modelos de predicción, porque permite calcular las probabilidades de que aparezcan datos y por ende poder inferir datos de la población a partir de los obtenidos de una muestra de la misma. Estas pruebas pueden ser las de *Shapiro-Wilk* para datos pequeños o *Kolmogorov-Smirnov* (con la corrección *Lillefors*) para datos más grandes. Si el *p-value* se es mayor a 0.05 se puede garantizar que los datos tienen una distribución normal.

Sin embargo existen datos que tienen una distribución sesgada ya sea a la derecha o a la izquierda que para su análisis requiere la aplicación de pruebas no paramétricas como las transformaciones logarítmicas y recíprocas (Orellana, 2008).

Las cuales permiten transformar las distribuciones ya sea comprimiendo los valores más altos y expandiendo los más pequeños balanceando la curva. La transformación se aplica directamente sobre cada valor en la ecuación.

2.3.2. Modelos predictivos

Un modelo es esquema o diseño en el que se representa una situación, proceso o entidad del mundo real, manteniendo cierta similitud y precisión respecto al mismo. En la actualidad los modelos son usados en distintas áreas para explicar una actividad realizando modificaciones sobre ellos cuando no es posible hacerlo directamente en la realidad.

Una predicción es una expresión en la que dadas ciertas condiciones se intenta anticipar a una situación. Los modelos predictivos o de predicción son aquellos que a partir de resultados anteriores evalúan la probabilidad de que suceda o no un evento, de esta manera se pueden simular distintos escenarios y obtener un pronóstico variando las condiciones iniciales en el modelo.

Para desarrollar un modelo de predicción primero se debe entrenar el sistema usando datos que se encuentren almacenados y contengan información del comportamiento de las personas, del tema o fenómeno a tratar. Estos datos deberán contener características propias que van variando de acuerdo a la disciplina aplicada. Una vez se obtienen, los datos pueden arrojar información sobre el comportamiento del fenómeno a partir de la observación.

Dicho de otra manera, los modelos predictivos dependen de variables, que según los valores ingresados modifican los resultados futuros. Para poder determinar las variables más relevantes del modelo se debe recopilar los datos del fenómeno a estudiar, una vez obtenidas se deben procesar a través de diferentes técnicas ya sean estadísticas o tecnológicas para obtener las predicciones.

En los últimos años los modelos predictivos han extendido sus aplicaciones y uso a las herramientas tecnológicas en donde se “enseña” a los programas a través de los datos para que se alimenten de patrones de esas experiencias y puedan tomar decisiones futuras. Este concepto se conoce como aprendizaje automático o *machine learning*, y para Briega (2015) es una disciplina que usa la estadística, la matemática y la computación para inducir una regla desconocida a partir de ejemplos donde esa regla es aplicada. A continuación se describe el algoritmo de

regresión lineal simple que es utilizado para resolver problemas de *machine learning*:

2.3.3. Modelos de regresión lineal simple

Es una técnica que permite determinar la relación entre variables. La regresión lineal se realiza cuando se desea saber la relación que existe entre una variable dependiente Y y una variable independiente X, además de proporcionar un balance sobre la eficacia del modelo a través del estudio de coeficientes de relación, análisis de residuos o varianzas (Pauly, 1983).

Al tener un conjunto de datos dispersos que muestran una tendencia creciente o decreciente y rectilínea con lo que se espera plantear una ecuación que permita encontrar el mejor ajuste de esos datos, para ello se usa la fórmula (Pauly, 1983):

Ecuación 4: Regresión Lineal

$$y = b + mx$$

Fuente: <http://www.fao.org/docrep/003/X6845S/X6845S00.HTM>

Para encontrar el cálculo de la pendiente m:

Ecuación 5: Cálculo de la pendiente

$$m = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Fuente: <http://www.fao.org/docrep/003/X6845S/X6845S00.HTM>

Mientras que el intercepto b se calcula:

Ecuación 6: Cálculo del intercepto

$$b = \left[\frac{\sum y}{n} - \left(m * \frac{\sum x}{n} \right) \right]$$

Fuente: <http://www.fao.org/docrep/003/X6845S/X6845S00.HTM>

Obteniendo los valores del intercepto a y de la pendiente b se puede obtener una línea recta que mejor se ajusta a los puntos dispersos de la muestra o datos experimentales. Gráficamente se puede confirmar si la recta representa el comportamiento de los puntos, si estos se acercan a la misma.

Cuando se aplica un método de regresión lineal se hace necesario realizar un análisis de correlación, el cual expresa la asociación entre dos variables representado por el coeficiente de correlación identificado como r . Este coeficiente puede tener valores que varían entre -1 y $+1$. El valor negativo en el coeficiente de correlación implica que una variable, la dependiente o la independiente tienden a decrecer cuando la otra aumenta y se le conoce como correlación negativa. Cuando es positivo, en cambio, esto significa que una variable se incrementa al hacerse mayor la otra (Pauly, 1983).

El coeficiente de correlación se calcula con la siguiente fórmula:

Ecuación 7: Coeficiente de determinación

$$r^2 = \frac{\sum xy - \left[\frac{(\sum x) (\sum y)}{n} \right]^2}{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] * \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}$$

Fuente: <http://www.fao.org/docrep/003/X6845S/X6845S00.HTM>

Ecuación 8: Coeficiente de correlación

$$|r| = \sqrt{r^2}$$

Fuente: <http://www.fao.org/docrep/003/X6845S/X6845S00.HTM>

2.4. SOFTWARE Y MODELOS DE PREDICCIÓN EXISTENTES APLICADOS A DEPORTES

2.4.1. Aplicación de un modelo matemático al *taper* en jóvenes nadadores

Este estudio pretende mostrar la aplicación de un modelo matemático a la reducción progresiva no lineal del nivel de entrenamiento en jóvenes nadadores antes de la competición. Se estudió el resultado de diez nadadores con pruebas antes y después del entrenamiento. Se tuvieron en cuenta las variables de volumen de

entrenamiento, la relación entre volumen e intensidad media de entrenamiento en una temporada (MITS) y la frecuencia de entrenamiento.

Este estudio duró catorce semanas, el volumen de nado fue de 8.200 mts diarios equivalentes a 47.000 *mts* semanales y la relación entre el volumen y la intensidad se controló con el índice *Mean Intensity Training Session (MITS)*. El índice se calcula de acuerdo a la fórmula que relaciona el volumen de entrenamiento completado a cada nivel de intensidad con su correspondiente producción de lactato, esta fórmula fue planteada por Mujika.

El modelo matemático consiste en un descenso exponencial lento de la carga de entrenamiento con una constante de declive de $8d^2$. En donde se relaciona el volumen e intensidad media del entrenamiento calculada para el día, el comienzo del *taper* y la constante de tiempo de descenso exponencial del *taper*.

En los resultados presentados en la investigación se puede determinar que las variables anteriormente descritas disminuyeron considerablemente. El modelo demuestra la correlación entre lactato máximo y marca, pero no muestra conclusiones irrefutables pues el estudio fue realizado sin grupo control por lo que pueden influir otras variables en el *taper* que pueden modificar la eficiencia del método.

Para Bisano (2014) “El período del *taper* está caracterizado por una reducción en la carga de entrenamiento que le permite al atleta recuperarse más rápido”. Las limitaciones del *taper* es que debe diseñarse de manera individual debido a las condiciones particulares de cada atleta.

2.4.2. Numerical Algorithms for Predicting Sports Results

En esta propuesta se plantea modelos matemáticos para predecir el resultado de partidos de fútbol americano, utilizando regresión logística. Se tiene en cuenta si el equipo es local o visitante y el modelo debe devolver una salida binaria indicando cuál de ellos tiene más probabilidades de ganar. En este estudio se pretende demostrar que el modelo planteado tiene igual éxito que los realizados en otras investigaciones, además se realiza una comparación del modelo con los resultados publicados en las casas de apuestas.

Principalmente se supone como primera medida que los equipos locales tienen cierta ventaja sobre los equipos visitantes, segundo se tienen en cuenta el marcador reciente de los equipos en partidos anteriores, y por último se aplica la regresión logística la cual tiene en cuenta variables como dotación y formación del equipo entre otros, esto con el fin analizar los atributos más importantes de cada partido indicando si se mantienen o no dependiendo del rendimiento de cada equipo.

2.4.3. A Hybrid Prediction System for American NFL Results

Se presenta un modelo para predecir los resultados de *NFL (National Football League)*, la base es la regresión lineal y el modelo de cadenas de Markov.

El modelo de Markov se usa para la clasificación de los equipos y se implementa una cadena con un estado para cada equipo. La intuición es un estado de transición simulando el comportamiento hipotético de un apostador en los sondeos o en las encuestas de las apuestas. El valor que toma el estado es el nombre del equipo el cual el apostador cree es el mejor, a medida que pasa el tiempo el apostador elige al azar un juego y el equipo el cual cree va ser el mejor contra su oponente y de acuerdo a la probabilidad (p) se va moviendo al equipo ganador y a una probabilidad ($1-p$) al equipo perdedor.

El sistema utiliza un enfoque de regresión lineal múltiple que implica que más de una variable de predicción está disponible y los componentes lineales representan los coeficientes de regresión que deben ser adicionados. El modelo investiga la relación lineal que existe entre las variables y los conjuntos de datos.

2.4.4. Sistema de predicción de resultados en eventos deportivos y su aplicación en las apuestas

Se presenta un sistema de predicción basado en clasificadores a partir de un volumen alto de información de partidos deportivos. Con técnicas de minería de datos se determina cuales atributos mejorarán la predicción, los datos son clasificados e ingresados a la herramienta WEKA que permitirá crear diversos modelos de predicción a partir de la clasificación que se les dio anteriormente. Se hace uso de Excel para automatizar la predicción de los resultados, a través del ingreso de los clasificadores a las macros se calcula el resultado más probable del partido y el riesgo estimado.

WEKA es una herramienta para análisis de datos que sirve para tratar los datos y clasificarlos de manera que queden homogéneos o se puedan clasificar cuales serán útiles para un estudio o investigación.

3. ESTRATEGIA METODOLÓGICA

3.1. MÉTODO DE INVESTIGACIÓN

La metodología desarrollada en el proyecto permitió la realización de un sistema orientado a crear la relación entre los datos y el proceso de análisis a los que serán sometidos, como en todo proyecto se hace necesario hacer uso de diferentes metodologías que den un contexto más claro sobre las herramientas y conceptos existentes para cumplir el objetivo. En primera instancia se determinó que esta investigación era del tipo experimental, de acuerdo a (Mantilla, 2002) se desarrolla un nuevo producto a partir de los conocimientos que se obtuvieron durante el desarrollo del proyecto.

También se hace uso de las herramientas conceptuales que ofrece la investigación descriptiva para poder tener un conocimiento más amplio de los temas directamente relacionados con el manejo y análisis de los datos además de la temática específica de natación. La investigación descriptiva tiene una ventaja adicional la cual se adapta perfectamente al desarrollo de actividades en este proyecto pues no sólo facilita la recolección de datos, sino que busca predecir e identificar la relación de las variables (Gross, 2014). Es así como también se apoyó el desarrollo en la investigación documental en cuanto hace uso de la recolección, selección, análisis y presentación de información para la formación de conocimiento que permitió la implementación de la aplicación web para el análisis del rendimiento en nadadores colombianos.

3.2. DESCRIPCIÓN DE LA METODOLOGÍA

Se plantearon las siguientes actividades: una fase inicial en donde se recolectó la información de distintas pruebas desarrolladas por expertos en el tema a diferentes atletas especializados en natación y datos históricos con los que se cuentan para dicha disciplina, posteriormente se procesó esta información, con el fin de obtener las variables más relevantes y concluir cuales se involucrarán en el modelo. Para procesar las variables y obtener resultados se debe hacer uso de herramientas estadísticas que permita determinar el comportamiento de las mismas.

Una vez se tengan definidas las variables y el modelo de predicción se implementará un prototipo haciendo uso del *framework Django y Bootstrap* con el cual se analizarán los resultados devueltos por el modelo. Al terminar se deberán

realizar las conclusiones a partir de los resultados arrojados por el modelo y ajustar el prototipo a recomendaciones que puedan ser de ayuda en el sistema.

Para la implementación, al ser un proyecto pequeño, se usó la metodología *XP* o *Extreme Programming*, por la facilidad de mezclarse con otras metodologías, adaptarse a las necesidades de desarrollo, de fácil documentación y por su propósito de desarrollar lo más rápido posible respetando los requerimientos manteniendo un contexto de simplicidad (Letelier & Penadés, 2016).

Las actividades se presentan divididas en 4 fases principales que se describen a continuación:

3.2.1. Primera fase: Generación de la propuesta

Contextualización:

En una primera medida se realizó un acercamiento a la problemática existente, con el propósito de tener un acercamiento al contexto inicial en el que se iba a desarrollar el proyecto. Posteriormente se pudo analizar la información que se tenía disponible y lo que se esperaba realizar en el proyecto a través de la evaluación de la situación y de esta manera obtener más claramente los objetivos, terminologías y conceptos propios del área.

Obtención del conocimiento:

Esta fase se desarrolló obteniendo todo el conocimiento necesario existente en cuanto a fundamentación, bases teóricas y técnicas en dos temas específicos, el rendimiento en natación y modelos predictivos y estrategias para su planteamiento. Se realizó consultas a diferentes libros y recursos electrónicos que permitieran dar una conceptualización más detallada y especializada de los temas. Posteriormente se realiza un inventario de las herramientas existentes que apliquen modelos de predicción en diferentes áreas y específicamente en deportes, esto con el objetivo de hacerse a una idea de cómo se puede podría abordar el problema.

La estrategia que se planteó fue hacer un registro sobre los conceptos y términos que se buscaban, tanto en español como en inglés, así como sinónimos o características que pudieran tener relación de los mismos. También como tener claridad a las fuentes que se podían acceder, primero a revistas especializadas en deportes, libros y páginas web que permitieran dar claridad a conceptos relacionados con natación, después a modelos predictivos existentes cuya fuente

principal fue el internet y las bases de datos de revistas especializadas, y por último a libros y recursos web relacionados con la teoría propia de modelos de regresión.

A medida que se fue desarrollando la estrategia, las fuentes consultadas fueron registradas en fichas bibliográficas haciendo uso del recurso web *¡Arma tu ficha Bibliográfica!* El cual permite el ingreso de bibliografía asociada a cualquier documento regulado bajo las normas APA, como se muestra en la

Figura 1.

Figura 1. Ficha bibliográfica

Fuente: (Universidad de Cuernavaca, s.f.)

Generación de la propuesta:

Se obtiene el documento con la propuesta de investigación en donde se detalla la descripción del problema, los objetivos, las referencias bibliográficas existentes respecto al tema, los resultados esperados y que sirvió de apoyo para completar la parte inicial del presente documento.

3.2.2. Segunda fase: Recopilación, análisis y comprensión de los datos

Se realizó la toma de datos teniendo en cuenta las fuentes principales de los mismos, se determinó la estrategia de cómo estos datos sería almacenados y tratados para su posterior análisis. En esta fase se realizaron actividades que se describen detalladamente en el RESULTADOS DE LA INVESTIGACIÓN

sección 4.2 del presente documento:

- Recopilación y descripción de los datos, página 26.

- Exploración de los datos, que información se tenía disponible y sería de utilidad para el desarrollo del modelo, página 27.
- Carga, visualización y resumen de los datos en R, página 29.

3.2.3. Tercera fase: diseño del modelo

Para esta fase se realizaron las siguientes actividades:

Basados en la bibliografía consultada se empezó a analizar cuál sería la mejor estrategia para el planteamiento del modelo aplicada a la contextualización del problema y los datos recopilados en las fases anteriores. Se procede a la construcción y evaluación del modelo haciendo uso del paquete estadístico *r*, en donde a medida que va devolviendo resultados y partiendo de los análisis estadísticos requeridos se seleccionan cuales datos arrojan información relevante en el modelo. Los resultados obtenidos se registraron en la sección 4.3 TERCERA FASE: SELECCIÓN DE LA TÉCNICA DE MODELADO

Análisis de datos con R. de los RESULTADOS DE LA INVESTIGACIÓN. Esta fase se desarrolla aplicando una metodología para la construcción de modelos la cual implica:

- Análisis de datos con R
- Diseño y ajuste del modelo

3.2.4. Cuarta fase: Implementación y despliegue

Se realizó una actividad preliminar de preparación del ambiente que consistió en realizar todas las configuraciones para poder dar inicio a la implementación del prototipo. También se instalaron las herramientas de desarrollo tales como el *Framework Django*, el lenguaje de programación *Python*, y el *Framework Bootstrap*.

La implementación se realizó usando la metodología XP en la que se especifican las siguientes fases:

- Definición de requerimientos: haciendo uso de las historias de usuario se determinaron cuáles serían los requerimientos a implementar.
- Diseño de Prototipo: para poder desarrollar esta actividad se usó la herramienta web *Cacoo* en la que se puede plasmar el diseño de las interfaces o *templates* con cada una de sus funcionalidades.
- Implementación: primero se realizaron los *templates* con el *Framework de JavaScript Bootstrap* para luego implementar a través de python las vistas o controladores de los *templates*.
- Pruebas: se tomaron 3 nadadores con diferentes cantidades de datos registradas y se realizaron estimaciones en distancias de 1500, 800 y 400 metros con el fin de determinar si los valores obtenidos eran similares a los esperados.

El resultado se detalla en la sección 4.4 CUARTA FASE: IMPLEMENTACIÓN DEL MODELO de los RESULTADOS DE LA INVESTIGACIÓN

4. RESULTADOS DE LA INVESTIGACIÓN

4.1. PRIMERA FASE: GENERACIÓN DE LA PROPUESTA

La primera fase del presente proyecto corresponde a todo el proceso de investigación de bibliografía y contexto que permitió obtener la propuesta de investigación y formar los aspectos generales y marco teórico registrado en este proyecto.

4.2. SEGUNDA FASE: RECOPIACIÓN, DESCRIPCIÓN, EXPLORACIÓN Y PREPARACIÓN DE LOS DATOS

4.2.1. Recopilación y descripción de los datos

Los datos usados son los que se encuentran disponibles en el histórico de la Federación Colombiana de Natación (FECNA) para nadadores colombianos. En un primer acercamiento la información se genera a partir de los filtros aplicados ya sea por deportista, evento y fecha de presentación de la prueba.

El acceso se realizó directamente a la página web y se exportaron los datos de cada deportista aplicando un único filtro a la fecha con rango de 01 de enero de 2000 a 31 de diciembre de 2016.

De cada nadador se extrajo información de nombre, fecha de nacimiento, evento, piscina, tiempo y la edad. En la *Figura 2* se presenta un reporte de la información obtenida:

Figura 2. Información almacenada en FECNA

Evento	Tiempo	Puntos Fina	Gan	Fecha	Lugar
50m Libre	00:28.04	204	16	2013-08-01	Med. 11
100m Libre	00:59.07	309	17	2013-08-01	Med. 11
200m Libre	01:37.81	202	11	2013-08-01	Med. 11
400m Libre	03:00.92	242	16	2013-08-01	Med. 11
800m Libre	05:07.10	309	15	2013-08-01	Med. 11
1500m Libre	09:52.41	310	17	2013-08-01	Med. 11
50m Estilo	01:02.21	225	15	2013-08-01	Med. 11
100m Estilo	02:02.00	186	10	2013-08-01	Med. 11
200m Estilo	04:09.01	226	16	2013-08-01	Med. 11
400m Estilo	08:00.12	310	17	2013-08-01	Med. 11
800m Estilo	16:07.24	315	18	2013-08-01	Med. 11
1500m Estilo	03:01.17	211	11	2013-08-01	Med. 11
100m Braza	01:54.00	174	10	2013-08-01	Med. 11

Fuente:

<http://www.fecna.com/index.php/natacion-carreras/estadisticas-carreras/historial-deportivo>

Cada dato se fue almacenando teniendo en cuenta:

- Información general: fecha de nacimiento, liga, club, piscina.
- Información demográfica del atleta: sexo, edad, ciudad, departamento.
- Información de la prueba: evento, tiempo, puntos obtenidos.

4.2.2. Exploración de los datos

Se recolectaron 89971 datos, pero esta información no se encontraba almacenada de la manera adecuada para iniciar el análisis, el primer paso fue descargarla uno a uno la información del deportista creando una base de datos que después se organizó de acuerdo a dos criterios: el evento realizado fue el de estilo libre, 50, 100, 200, 400, 800 y 1500 metros, y la edad del nadador en el momento de la prueba, rango de 11 a 17 años.

Cruzando la información de edad, distancia y sexo se obtienen los datos de cada deportista como se muestran en las

Tabla 1 y Tabla 2:

Tabla 1 - Datos recolectados para hombres

HOMBRE = 1					
Edad vs Distancia	50	100	200	400	800
11	907	680	111	656	10
12	1113	1053	114	890	253
13	953	947	541	701	276
14	991	1006	772	642	282
15	800	826	643	518	259
16	559	576	461	372	164
17	349	382	289	227	116
Total	5722	5570	3131	4406	2160
	20989				

Fuente: elaboración propia

Tabla 2 - Datos recolectados para mujeres

MUJER = 2					
Edad vs Distancia	50	100	200	400	800
11	702	625	79	501	113
12	864	813	385	535	241
13	823	807	558	473	225
14	712	663	537	391	231
15	466	483	367	261	183
16	300	298	270	183	111
17	160	185	130	112	84
Total	4077	3974	2526	2856	1988
	15421				

Fuente: elaboración propia

El objetivo será obtener conocer el comportamiento de los datos explorando más profundamente en la clasificación de los mismos.

4.3. TERCERA FASE: SELECCIÓN DE LA TÉCNICA DE MODELADO

4.3.1. Análisis de datos con R.

Los datos se encuentran almacenados en un archivo plano llamado datos, lo que se espera es visualizar el comportamiento de variables tiempo, distancia, resultados de acuerdo a la edad, la liga, la prueba, el lugar, el género y su influencia en el tiempo de ejecución de una prueba.

Estos datos fueron cargados en R creando un objeto llamado “muestra” e importándolo así:

```
datos<-read.table("muestra.txt",header=T,sep="\t",dec="," ,na.strings="NA")
```

El encabezado contiene los títulos de las columnas por eso el *header* es T, Los datos decimales en él .txt se encuentran separados por ‘,’ , *na.string* se refiere a los valores desconocidos que por defecto para R el valor es NA.

A través del comando *fix(datos)* se puede visualizar la estructura general de los datos y de esta manera garantizar que han sido cargados adecuadamente, tal como se muestra en la Figura 3.

Figura 3. Carga de todos los datos y variables

ID	GENERO	NOMBRE	FECHA_NACIMIENTO	Ciudad	DISTANCIA	ESTILO	TIEMPO	PUNTOS	EDAD	FECHA
1000001	M	ANTHONY	1998-07-08	BOGOTA	400	Libre	33.55	0	11:22	2001-06-31
1000002	F	SOFIA	1999-11-18	VALLÉ	505	Libre	182.13	396.3	13:28	2011-03-31
1000003	M	ANTHONY	1998-07-23	ASTIGUIA	381	Libre	236.06	426.5	13:30	2016-03-18
1000004	M	ANTHONY	1998-01-28	SANTANDER	103	Libre	441.11	1429.1	14:30	2012-09-27
1000005	M	ANTHONY	1998-07-04	TOLIMA	33	Libre	620.40	1531.0	15:30	2015-05-30
1000006	M	ANTHONY	1998-03-30	VALLÉ	35	Libre	4214.14	1714.0	16:30	2016-07-04

Fuente: elaboración propia

Para la determinación del modelo se debe tener en cuenta el análisis del comportamiento de los datos, de manera que se puede observar gráfica y estadísticamente la correlación del tiempo, la edad, la prueba y los puntos obtenidos. Para este estudio se hizo uso del paquete r.

Se obtiene un resumen numérico de los datos a través del comando *summary* tal como se muestra en la Figura 4:

Figura 4. Resumen estadístico arrojado por R

```

> summary(datos)
      GÉNERO      NACIMIENTO      LUGAR      DISTANCIA      ESTILO      TIEMPO      PUNTOS      EDAD      FECHA
NOMBRE: 364 Min. :1997-07-08 00:00:00 BOGOTÁ :403 Min. : 100.0 Libres:1395 Min. : 33.55 Min. : 0.0 11:22 Min. :2001-06-31 00:00:00
      1st Qu.:1896-11-18 00:00:00 VALLÉ :505 1st Qu.: 200.0 1st Qu.: 182.13 1st Qu.:396.3 13:28 1st Qu.:2011-03-31 00:00:00
      Median :1998-07-23 00:00:00 ASTIGUIA:381 Median : 400.0 Median : 236.06 Median :426.5 13:30 Median :2016-03-18 00:00:00
      Mean :1998-01-28 14:40:00 SANTANDER:103 Mean : 341.1 Mean : 441.11 Mean :1429.1 14:30 Mean :2012-09-27 02:51:31
      3rd Qu.:12000-07-04 12:00:00 TOLIMA : 33 3rd Qu.: 800.0 3rd Qu.: 620.40 3rd Qu.:1531.0 15:30 3rd Qu.:2015-05-30 00:00:00
      Max. :2008-03-30 00:00:00 VALLÉ : 35 Max. :1500.0 Max. : 4214.14 Max. :1714.0 16:30 Max. :2016-07-04 00:00:00
      (Other) :284
      CIUDAD      DEPARTAMENTO      #<100.50      #<200.100      #<400.200      #<800.400      #<1500.800
Val :580 Cali :520 2.20000000000000002 : 32 Min. : 11.600 Min. : 0.390 Min. : 10.170 Min. : 11.480
Seg :340 DC : 230 1.16 : 30 1st Qu.:12.045 1st Qu.: 1.980 1st Qu.:11.810 1st Qu.:11.820
Ant :408 Medellín:328 2.06 : 20 Median :12.190 Median : 2.120 Median :12.030 Median :11.930
Cun :135 Ciudadot:135 2.13 : 17 Mean :12.205 Mean : 2.079 Mean :12.026 Mean :11.911
Sia :128 Pereira :128 2.12 : 17 3rd Qu.:12.380 3rd Qu.: 2.230 3rd Qu.:12.160 3rd Qu.:12.020
San : 88 (Other) :332 (Other) : 203 Max. :13.150 Max. : 23.870 Max. :12.850 Max. :12.460
(Other):410 NA's : 6 NA's :1378 NA's :1563 NA's :1609 NA's :1670
    
```

Fuente: elaboración propia

Se generó la matriz de correlación (Figura 5) para poder saber si existe relación entre el tiempo, la edad, el sexo, distancia y puntos.

Figura 5. Análisis de correlación de los datos

```
> cor(Dataset[,c("EDAD", "PRUEBA", "PUNTOS.", "SEXO", "TIEMPO")], use="complete")
      EDAD      PRUEBA      PUNTOS.      SEXO      TIEMPO
EDAD    1.000000000  0.118671444  0.60384107 -0.163904940  0.009118946
PRUEBA  0.118671444  1.000000000  0.14552664 -0.040080632  0.004606741
PUNTOS. 0.603841071  0.145526642  1.00000000  0.070556296 -0.021428436
SEXO   -0.163904940 -0.040080632  0.07055630  1.000000000 -0.005799648
TIEMPO  0.009118946  0.004606741 -0.02142844 -0.005799648  1.000000000
```

Fuente: elaboración propia

En la correlación se verifica aquellos valores que más se acercan a 1, por ejemplo existe cierta correlación entre el tiempo, la distancia y la edad, pero en menor medida entre el tiempo, los puntos y el sexo por lo que serán dos variables que se pueden eliminar del modelo.

Ahora se analiza la distribución normal de los datos con una prueba de *Lilliefors* donde el *p-value* debe ser mayor a 0.05 para que los datos tengan una distribución normal, de acuerdo a los resultados este valor es muy pequeño lo que da una idea de la poca normalidad de los datos como se muestra en la Figura 6.

Figura 6. Estudio de la distribución normal a partir de *Lilliefors*.

```
> with(Dataset, lillie.test(PRUEBA))
      Lilliefors (Kolmogorov-Smirnov) normality test
data:  PRUEBA
D = 0.28505, p-value < 2.2e-16

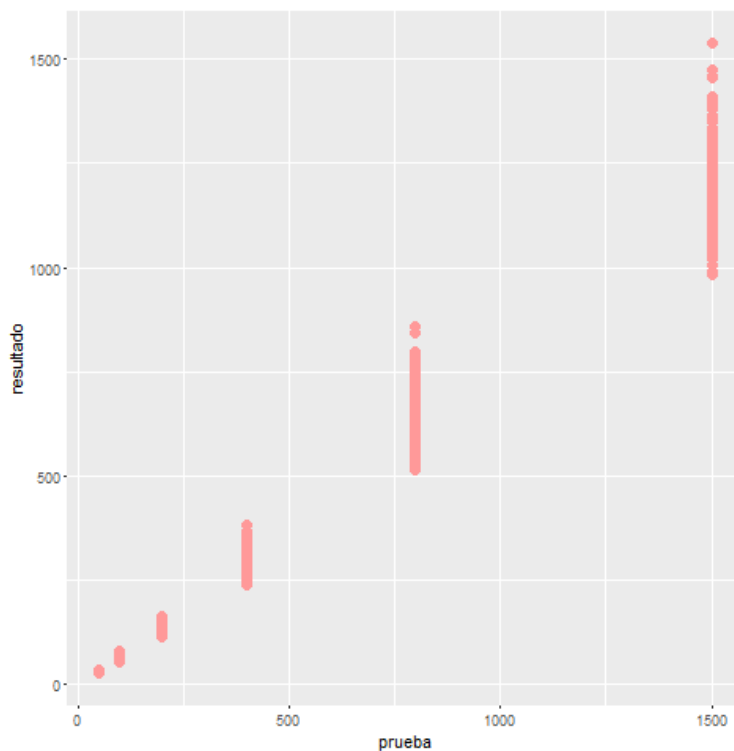
> with(Dataset, lillie.test(PUNTOS.))
      Lilliefors (Kolmogorov-Smirnov) normality test
data:  PUNTOS.
D = 0.015878, p-value = 1.284e-13

> with(Dataset, lillie.test(TIEMPO))
      Lilliefors (Kolmogorov-Smirnov) normality test
data:  TIEMPO
D = 0.49466, p-value < 2.2e-16
```

Fuente: elaboración propia

La Figura 7 permite analizar de manera gráfica el comportamiento entre la distancia y el tiempo, confirmando que cargando todos los datos y realizando un análisis general, estos no muestran una distribución normal porque el comportamiento no es acampanado o no hay una curva que pueda ajustarse al mismo. Sin embargo en la gráfica se puede observar que los datos se agrupan de acuerdo a la distancia, por lo que se genera la hipótesis que probablemente especificando el análisis por rangos de distancia de 100, 200, 400, 800 y 1500 metros se pueda validar si existe o no la normalidad.

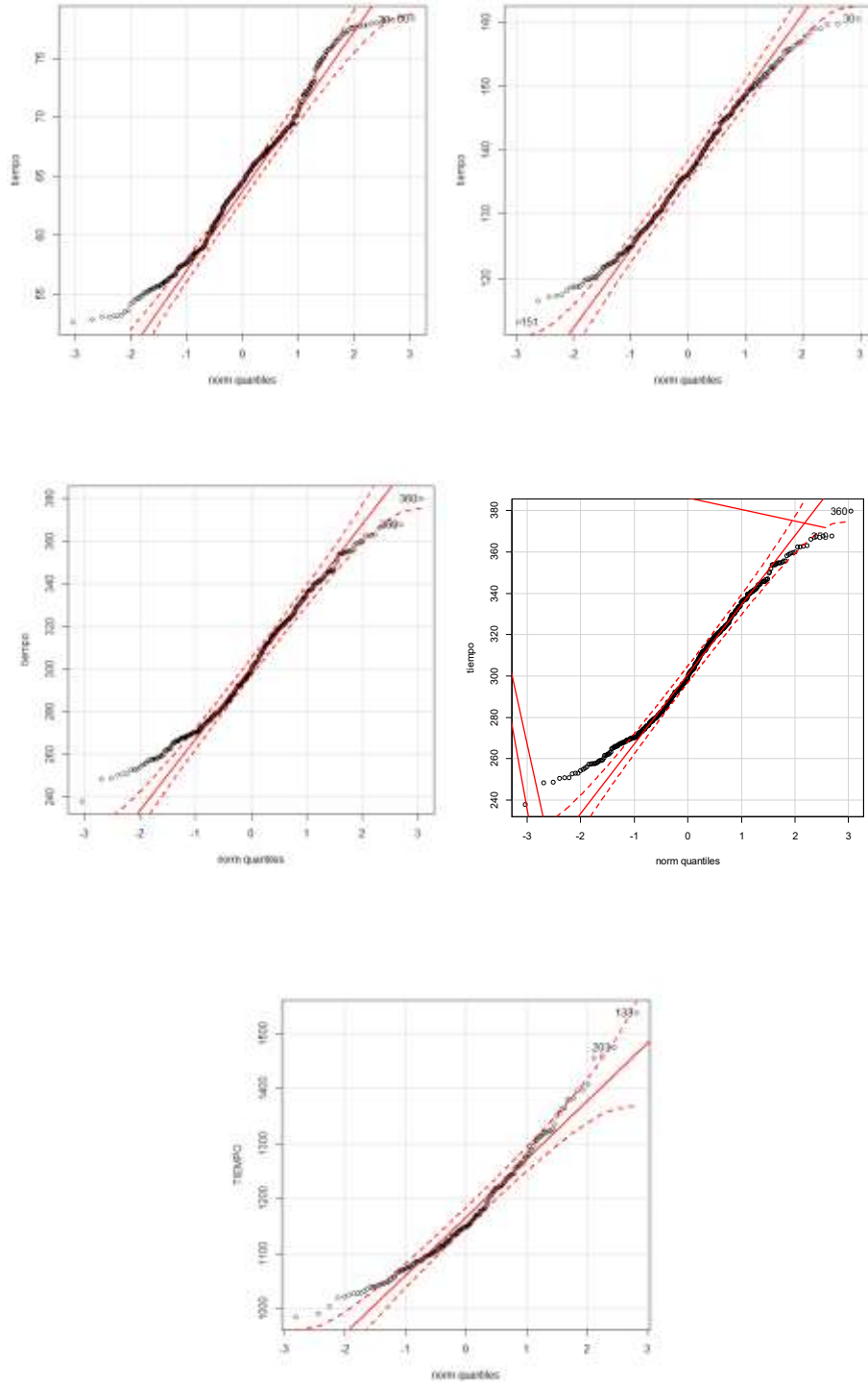
Figura 7. Comportamiento del tiempo (resultado) vs la prueba (distancia).



Fuente: elaboración propia

Con las gráficas de normalidad (Figura 8) se verifica que en cada distancia, el tiempo tiene una distribución exponencial, es decir tiende a crecer en cuantos más datos se tengan, por lo que pueden ajustarse a través de modelos de regresión a la dispersión de los puntos:

Figura 8. Gráficas de normalidad del tiempo en distancias 100, 200, 400, 800 y 1500 mts.

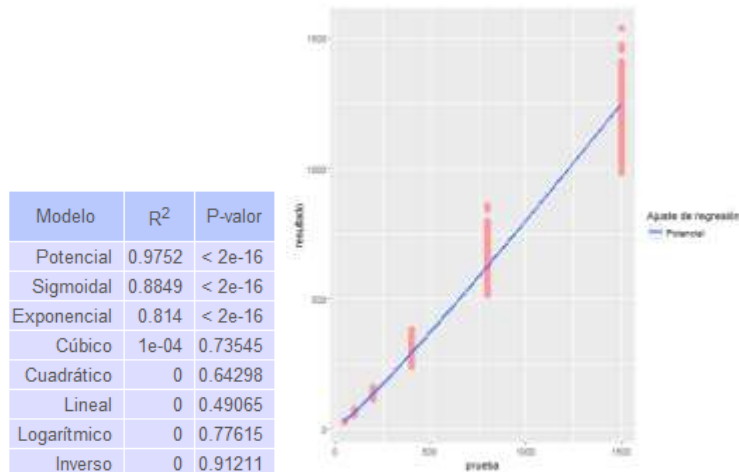


Fuente: elaboración propia

4.3.2. Diseño del modelo

El análisis de los datos permite concluir que no se puede implementar un modelo de regresión directamente sobre toda la muestra, se confirma con la *Figura 9* en donde se intenta ajustar una curva que simule el comportamiento de los datos. Sin embargo el estudio de la normalidad de la *Figura 8* y la *Figura 9* permiten indicar que existe un comportamiento creciente en los datos pero que estos deben ser implementados o aplicados en otras ecuaciones con el fin de garantizar que arrojen resultados válidos para la implementación del modelo.

Figura 9. Mejor modelo de regresión y gráfica adaptada al comportamiento de los datos.



Fuente: elaboración propia

Al hacer el estudio de la bibliografía se pudo determinar que el tiempo y la distancia aplicados a las ecuaciones de la velocidad crítica (CSS) y la resistencia (FF), entendidos como indicadores de rendimiento, permiten estimar un tiempo muy cercano al obtenido por los atletas dada una distancia específica. Para cada fórmula se hace necesario plantear un modelo de regresión lineal dado que ambas ecuaciones se miden a partir de dos valores de distancia y tiempo; el objetivo será hacer los cálculos usando grandes volúmenes de datos que permitan ajustar la estimación al valor real.

De acuerdo a la Ecuación 2 el cálculo de la distancia depende del tiempo, la capacidad anaeróbica (ASC) y la CSS. Como lo que se espera es que el modelo

estime el tiempo que un nadador toma en recorrer una distancia objetivo, se realiza el despeje del tiempo en la ecuación de manera que:

Ecuación 9: Cálculo del tiempo a partir del CSS y ASC

$$\text{Tiempo} = (\text{distancia} - \text{ASC}) / \text{CSS}$$

Fuente: elaboración propia

El siguiente paso entonces es encontrar la forma de calcular la CSS y ASC a través de la regresión lineal usando los datos que se capturaron y analizaron en la fase anterior. Se hace uso de la **¡Error! No se encuentra el origen de la referencia.** para calcular la pendiente m que representa la CSS y la Ecuación 6 permite obtener el intercepto b que representa la capacidad anaeróbica (ASC).

Para calcular la ecuación de resistencia planteada por (Riegel, 1977) en *Athletic Records and Human Endurance*.y se tiene que:

Ecuación 10: Fórmula del cálculo del tiempo planteada por Riegel

$$\text{Tiempo estimado} = b * \text{distancia objetivo}^m$$

Fuente: Athletic Records and Human Endurance, American Scientist. American Scientist, 285-290

Los valores m y b se obtuvieron a través de la regresión lineal aplicada a la transformación de los datos a través del logaritmo natural de la distancia y el tiempo. El factor fatiga (FF) es el valor de la pendiente m , cuya relación el tiempo será inversamente proporcional es decir el tiempo será mayor a medida que aumente el factor fatiga y este se incrementará el doble a medida que se incremente la distancia. El valor del intercepto b será quien determine el comportamiento de la curva.

En el caso de FF se realiza el ajuste en la fórmula incluyendo la transformación logarítmica así:

Ecuación 11: Cálculo de la pendiente aplicando transformación de los datos

$$m = \frac{\sum \ln x \sum \ln y - n \sum (\ln x \ln y)}{(\sum \ln x)^2 - n \sum (\ln x)^2}$$

Fuente: elaboración propia

Para calcular el intercepto se tuvo que realizar la operación contraria al logaritmo de manera que la formula se modificó así:

Ecuación 12: Transformación del cálculo del intercepto

$$b' = e^b$$

Fuente: elaboración propia

Donde b es la fórmula del intercepto planteada en la Ecuación 6.

Se tomaron algunos datos de nadadores registrados en la FECNA y se fueron creando los modelos de acuerdo a las ecuaciones; estos resultados sirvieron para probar que el modelo implementado en la aplicación web daba resultados correctos.

Figura 10. Modelo implementado en R.

```
Call:
lm(formula = V2 ~ V3, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-57.731  -7.879  -0.405   5.499  36.898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.42675    9.69774   1.694   0.116
V3           0.95682    0.03945  24.252 1.45e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.35 on 12 degrees of freedom
Multiple R-squared:  0.98, Adjusted R-squared:  0.9783
F-statistic: 588.1 on 1 and 12 DF, p-value: 1.454e-11
```

Fuente: elaboración propia

4.3.3. Evaluación del modelo

Para la evaluación del modelo se tuvo en cuenta el cálculo de la precisión a través de la obtención del error porcentual medio absoluto (MAPE). El MAPE se calcula cuando se tienen modelos en los que se espera que la predicción se encuentre por debajo de un porcentaje de error, en otras palabras es un indicador del desempeño o efectividad del pronóstico. Este tipo de indicador no se debe usar cuando se tienen pequeños volúmenes de datos.

En el caso específico de esta investigación la precisión se entiende como el porcentaje en los cuales el modelo tuvo el resultado más acertado posible indicado por un valor del MAPE inferior al 20%. Se realizó la consulta de otros indicadores de rendimiento en modelos de predicción pero fueron descartados debido a que su aplicabilidad va directamente relacionada con modelos enfocados a clasificadores.

Para ello se deben tener en cuenta la resolución de las siguientes ecuaciones:

1. Suma acumulada de errores de pronóstico (CFE): Es la suma acumulada de los errores de pronóstico, el cual se calcula calculando la diferencia entre el valor real y el obtenido a través de la predicción.

Ecuación 13: Suma acumulada de errores de pronóstico

$$CFE = \sum \text{Error de pronóstico}$$

Fuente: <https://ingenioempresa.com/medicion-error-pronostico/>

2. Desviación media absoluta (MAD): Mide la dispersión o el tamaño del error del pronóstico en unidades. Es el valor absoluto de la diferencia entre la demanda real y el pronóstico, dividido sobre el número de registros. Como se muestra en la Ecuación 14.

Ecuación 14: Desviación media absoluta MAD

$$MAD = \frac{\sum |Real - Pronóstico|}{n}$$

Fuente: <https://ingenioempresa.com/medicion-error-pronostico/>

3. Error cuadrático medio (MSE): de acuerdo a (Ingenio Empresa, 2016) “es una medida de dispersión del error de pronóstico, sin embargo esta medida maximiza el error al elevar al cuadrado, castigando aquellos periodos donde la diferencia fue más alta a comparación de otros”.

Ecuación 15: Error cuadrático medio MSE

$$MSE = \frac{\sum \text{Error de pronóstico}^2}{n}$$

Fuente: <https://ingenioempresa.com/medicion-error-pronostico/>

1. Raíz del error cuadrático medio (RMSE): es la raíz cuadrada de la sumatoria de los errores cuadráticos. En comparación con la Desviación Media Absoluta o MAD, RMSE amplifica y penaliza con mayor fuerza aquellos errores de mayor magnitud.
2. Error porcentual medio absoluto (MAPE): es el porcentaje de la de la diferencia entre valor real y el pronóstico.

Ecuación 16: Error porcentual medio absoluto MAPE

$$MAPE = \frac{\sum_{i=1}^n 100 |Real_i - Pronóstico_i|}{Real_i \cdot n}$$

Fuente: <https://ingenioempresa.com/medicion-error-pronostico/>

La metodología usada para determinar la eficiencia del modelo fue separar los valores de entrenamiento de los valores para realizar la prueba de predicción, con muestras aleatorias con los siguientes porcentajes:

- 30% para entrenamiento y 70% para prueba
- 50% de entrenamiento y 50% de prueba
- 70% entrenamiento y 30% de prueba.

El resultado de la efectividad del modelo dada por MAPE se presenta en la Tabla 3

Tabla 3 - Cálculo de la precisión del modelo a través del MAPE con muestras aleatorias

Forecast Accuracy (Precisión de la predicción)												
	CFE con CSS	CFE con FF	MAD con CSS	MAD con FF	MSE con CSS	MSE con FF	RMSE CSS	RMSE FF	MAPE CSS	MAPE FF	PRECISIÓN CSS	PRECISIÓN FF
30-70	-286,13	-828,76	38,74	47,30	3085,85	5266,05	55,55	72,57	9,14%	8,58%	90,86%	91,42%
50-50	-317,27	-610,59	40,80	50,00	3263,84	5114,13	57,13	71,51	8,54%	9,14%	91,46%	90,86%
70-30	-133,91	-109,74	29,58	56,44	1757,89	4032,90	41,93	63,51	8,95%	23,5%	91,05%	76,50%
Promedio	-245,77	-516,36	36,37	51,24	2702,53	4804,36	51,54	69,20	9%	14%	91%	86%

Fuente: elaboración propia

De acuerdo al valor promediado del MAPE el error se encuentra por debajo del 20% por lo que las predicciones en ambos modelos son bastante acertadas. El modelo mejor ajustado es el de velocidad crítica, el cual a medida que se incrementan los valores de entrenamiento las predicciones son más precisas. Para el caso del modelo de factor fatiga se observa que tiende a disminuir la precisión cuando se incrementan los datos de entrenamiento esto se puede deber a la heterogeneidad de la muestra con la que fue entrenada el modelo, es por esta razón que con el fin de obtener mejores valores en la precisión de la estimación a partir del factor fatiga se realizó una nueva toma de muestras, prestando especial atención en que se mantuviera una uniformidad en la distribución de los datos dada la clasificación de las distancias, así se pudo garantizar que al realizar la estimación el modelo tenía suficientes valores para poder predecir correctamente el comportamiento dada una distancia específica. En los resultados se observa una mejora considerable en la precisión del modelo de factor fatiga como se muestra en la

Tabla 4.

Tabla 4 - Cálculo de la precisión del modelo a través del MAPE con muestras con distribución uniforme en la clasificación de las distancias

Muestras

Forecast Accuracy (Precisión de la predicción)

	CFE con CSS	CFE con FF	MAD con CSS	MAD con FF	MSE con CSS	MSE con FF	MAPE CSS	MAPE FF	PRECISIÓN CSS	PRECISIÓN FF
30-70	-1097,00	-828,76	23,37	47,30	1810,47	5266,05	6,36%	8,58%	93,6%	91,4%
50-50	-1209,92	-661,25	36,10	47,07	4399,40	5056,98	4,37%	6,80%	95,6%	93,2%
70-30	-133,91	739,18	29,58	26,99	1757,89	2520,24	3,92%	3,00%	96,1%	97,0%
TOTALES	-813,61	-250,28	29,68	40,45	2655,92	4281,09	4,88%	6,13%	95,12%	93,87%

Fuente: elaboración propia

4.4. CUARTA FASE: IMPLEMENTACIÓN DEL MODELO

4.4.1. Definición de requerimientos

Para la definición de requerimientos se hizo uso de las historias de usuario (HU), las HU son artefactos cortos y simples que permiten definir lo que el usuario necesita hacer, sin detallar el cómo (Morillo, 2015).

Las historias de usuario fueron desarrolladas a partir de la revisión de la bibliografía y del análisis obtenido en la etapa anterior, pues fue en esta que se pudo aclarar que datos se requerían para la implementación y que resultados debían mostrarse en el sistema para realizar el análisis de rendimiento por parte del usuario.

Para el registro de historias de usuario se utilizó el formato de la

Figura 11.

Figura 11. Plantilla para registro de HU.

Desarrollo ágil: Historias de usuario y criterios de aceptación
Elaborado por: www.gmoinformatica.com

Identificador (ID de la historia)	Rol	Escenario de la historia		Criterios de aceptación				
		Características / Funcionalidad	Pasos / Resultado	Número (ID de escenario)	Criterio de aceptación (Ited)	Condición	Resultado	Resultado / Comportamiento esperado

Fuente: (Morillo, 2015)

A partir de las historias de usuario se pudieron generar el listado de requerimientos los cuales se muestran en la Tabla 5.

Tabla 5 - Lista de requerimientos

Id. Req	Descripción	Tipo
1	La aplicación contará con un formulario para registro de pruebas solicitando se ingrese la prueba (distancia) y el tiempo.	Funcional
2	La página principal de la aplicación es el listado de pruebas registradas.	Funcional
3	En el listado de pruebas registradas se permite seleccionar algunas marcas o todas para realizar el análisis.	Funcional
4	En el listado de pruebas registradas aparecen botones para facilitar la navegabilidad como crear una prueba nueva o subir un archivo con varios resultados de prueba.	Presentación - Navegabilidad
5	A partir de la selección de algunas pruebas y del ingreso de una distancia objetivo, en la página de listados de pruebas se generan los resultados.	Funcional
6	En el reporte de resultados muestra valores calculados como los ritmos para cada pruebas, las estimaciones de acuerdo a la regresión utilizada así como los valores de los ritmos por nivel	Funcional
7	En el reporte de resultados se permite devolverse al listado de pruebas o al registro de las mismas.	Presentación - Navegabilidad
8	En las vistas existe un menú superior para acceder a las opciones para crear, listar o generar resultados.	Presentación - Navegabilidad
9	El listado de pruebas muestra opciones para eliminar el registro. Este se borra de la base de datos	Funcional

Tabla 5 (Continuación)

Id. Req	Descripción	Tipo
10	El listado de pruebas muestra opciones para editar el registro. Este se actualiza en la base de datos.	Funcional
11	El sistema permite guardar pruebas individualmente o a través de la carga masiva de registros	Funcional
12	El sistema tiene una sección de ayuda donde se encuentra información sobre los conceptos de natación usados en la aplicación e información de uso de la misma.	Funcional

Fuente: elaboración propia

A partir de las HU se pudo establecer que se implementarán las siguientes vistas:

- **Formulario de registro de pruebas:** es un formulario sencillo donde se registran los datos referentes a las pruebas, como son la distancia y el tiempo.
- **Listar pruebas:** es una tabla donde se listan las pruebas registradas. Permite elegir las pruebas que desea el nadador realizar el análisis y el ingreso de la distancia objetivo. Desde esta interfaz se debe poder acceder a los formularios de crear pruebas, cargar pruebas masivas y generar resultados.
- **Resultados:** una tabla con los resultados de aplicar los modelos de regresión, cálculos de ritmos por prueba y por nivel.
- **Manual de usuario:** Página donde se encontrará información de la aplicación, conceptualización general sobre velocidad crítica y ecuaciones de resistencia.

4.4.2. Diseño de Prototipo

Para la implementación se desarrollaron los prototipos de las interfaces a través del software <https://cacao.com>, la cual es una aplicación web y gratuita que permite realizar todo tipo de diagramas. La implementación de las interfaces se muestra en las siguientes ilustraciones:

Figura 12. Interfaz para listar pruebas



Fuente: elaboración propia

Figura 13. Formulario para registro y edición de prueba



Fuente: elaboración propia

Figura 14. Interfaz para confirmación de eliminación de registro



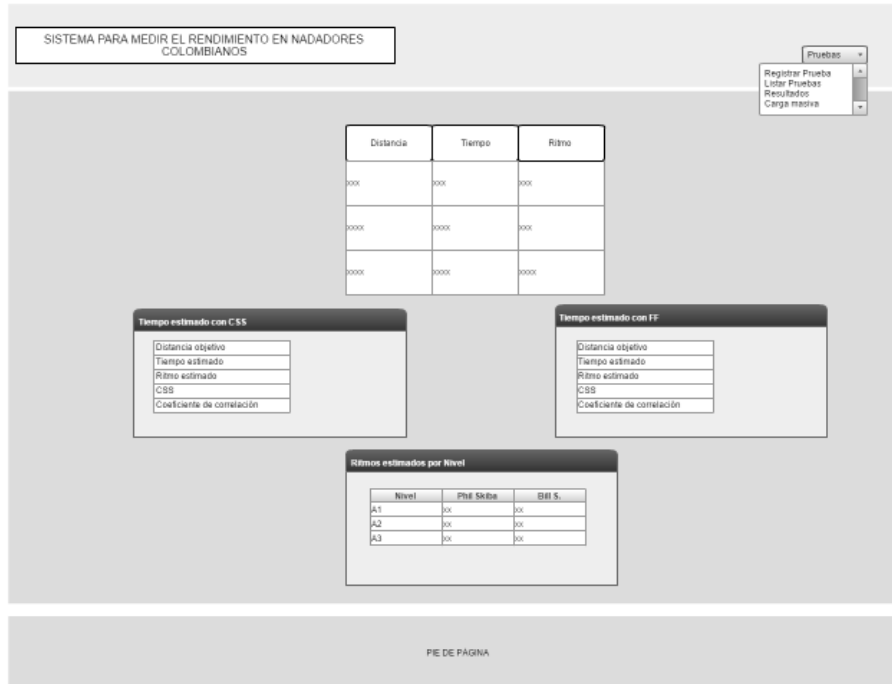
Fuente: elaboración propia

Figura 15. Interfaz para carga masiva



Fuente: elaboración propia

Figura 16. Interfaz para visualización de resultados



Fuente: elaboración propia

4.4.3. Implementación

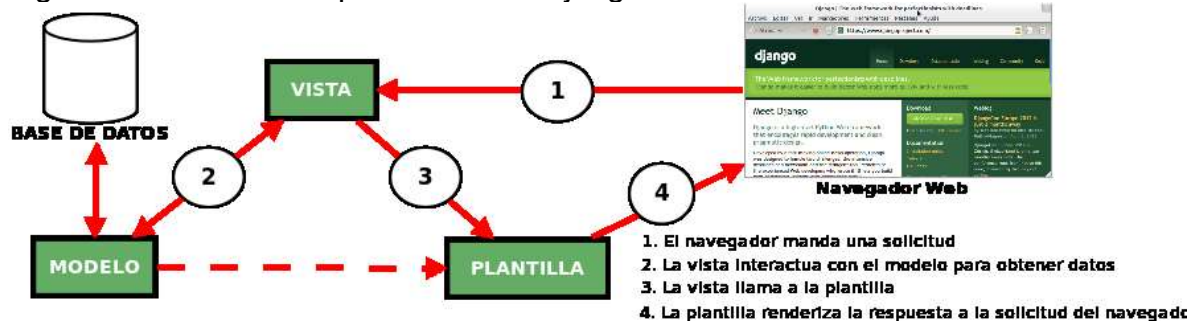
4.4.3.1. Selección de tecnologías:

Para las interfaces se usó el *framework* de *Javascript Bootstrap* que permite la realización de interfaces amigables y adaptables al dispositivo. Para la implementación del modelo se usó el lenguaje de programación *Python* a través del *Framework Django*, este *framework* es uno de los más usados para aplicaciones web dado que agiliza la implementación de cualquier proyecto por su filosofía de reutilización de código, permite escribir código robusto y fácil de entender al hacer uso de *Python*, un sistema de administración de usuarios incluido y facilidad de uso del sistema ORM para consulta en base de datos.

Django se basa en el modelo *template* (plantilla) vista o mejor conocido como MTV:

- “M significa *"Model"* (Modelo), la capa de acceso a la base de datos. Esta capa contiene toda la información sobre los datos: cómo acceder a estos, cómo validarlos, cuál es el comportamiento que tiene, y las relaciones entre los datos.
- T significa *"Template"* (Plantilla), la capa de presentación. Esta capa contiene las decisiones relacionadas a la presentación: como algunas cosas son mostradas sobre una página web u otro tipo de documento.
- V significa *"View"* (Vista), la capa de la lógica de negocios. Esta capa contiene la lógica que accede al modelo y la delega a la plantilla apropiada: ES como un puente entre el modelos y las plantillas” (Holovaty & Kaplan-Moss, 2015, pág. 90).

Figura 17. Modelo Template Vista de Django



Fuente:

<http://www.maestrosdelweb.com/curso-django-entendiendo-como-trabaja-django/>

4.4.3.2. Diseño de Base de datos

El modelo de base de datos se definió a través del *Framework Django* el cual cuenta con un mapeo objeto-relacional (ORM) que permite el acceso desde las vistas.

Fueron creadas tres tablas o modelos donde se almacenan los datos de cada prueba y los resultados que genera el modelo. Las tablas se identifican por tener la primera letra en mayúscula, mientras que los campos en cada tabla están nombrados en minúscula e intentan, a través del nombre, ofrecer una descripción del dato que almacenan.

En la *Figura 18*, se muestra la definición del modelo. Una vez definidos las tablas y campos se procede a actualizar la base de datos en *postgres*. Esta acción se puede realizar desde el mismo *framework* a través de las migraciones.

Figura 18. Modelo de base de datos.

```
from django.db import models

# Create your models here.

class Prueba(models.Model):
    distancia = models.FloatField()
    tiempo = models.FloatField()

class Resultado(models.Model):
    prueba = models.ForeignKey(Prueba, null=True, blank=True)
    distancia_objetivo = models.CharField(max_length=10)
    tiempo_estimado_ss = models.TimeField()
    tiempo_estimado_vm = models.TimeField()
    tiempo_estimado_ff = models.TimeField()
    ritmo = models.TimeField(null=True, blank=True)
    css = models.TimeField(null=True, blank=True)
    ff = models.FloatField(null=True)
    distancia_objetivo = models.FloatField(null=True)
    tiempo_estimado = models.TimeField(null=True, blank=True)
    ritmo_estimado = models.TimeField(null=True, blank=True)
    regenerativo = models.TimeField(null=True, blank=True)
    resistencia = models.TimeField(null=True, blank=True)
    tempo = models.TimeField(null=True, blank=True)
    umbral = models.TimeField(null=True, blank=True)
    vomax = models.TimeField(null=True, blank=True)
    distancia_objetivo_s = models.FloatField(null=True)
    capacidad = models.TimeField(null=True, blank=True)
    sprint = models.TimeField(null=True, blank=True)
    aerobico1 = models.TimeField(null=True, blank=True)
    aerobico2 = models.TimeField(null=True, blank=True)
    aerobico3 = models.TimeField(null=True, blank=True)
    umbral_bill = models.TimeField(null=True, blank=True)
    resistencia_alto_rendimiento = models.TimeField(null=True, blank=True)
    anaerobico = models.TimeField(null=True, blank=True)
    sprint_bill = models.TimeField(null=True, blank=True)
```

Fuente: elaboración propia

4.4.3.3. Diseño de plantillas o templates

Las interfaces o *templates* en la implementación fueron desarrolladas con el *framework* de *javascript Bootstrap* lo que permite que sean agradables a la vista y se adapten a cualquier dispositivo.

El software permite el ingreso de los datos de las pruebas que se desean analizar de dos formas: a través de un formulario como se muestra en la

Figura 19 o través de la carga masiva de registros almacenados en un archivo .csv tal como se indica en la Figura 20.

Cada valor ingresado sirve para darles precisión a los modelos de regresión en las fórmulas.

Figura 19. Formulario de ingreso de pruebas

Estimación del rendimiento
en Nadadores Colombianos

Pruebas y Resultados ▾

Formulario para el ingreso de pruebas

Distancia

Tiempo

Guardar Cancelar

ACERCA DE LA APLICACIÓN

Esta aplicación permite el ingreso de las marcas obtenidas por los nadadores en sus pruebas ya sean de competencia o en entrenamientos y realizar cálculos basados en modelos de regresión aplicados a la velocidad crítica y el factor fatiga. Estos modelos han sido entrenados con datos de la Federación Colombiana de Natación - FECNA e intentar dar un estimado del rendimiento del nadador dada una distancia objetivo.

ENLACES DE INTERÉS

HOME

FEDERACIÓN COLOMBIANA DE NATACION

Desarrollado por : Paula Figueroa

Fuente: elaboración propia

En la Figura 20 se visualiza la interfaz para la carga masiva de registros, de esta manera a través de un archivo .csv el usuario puede registrar sus registros en el aplicativo y así realizar los análisis posteriores.

Figura 20. Carga masiva de registros



Fuente: elaboración propia

Desde la aplicación se pueden eliminar registros como se muestra en la *Figura 21:*

Figura 21. Confirmación de eliminación de un registro



Fuente: elaboración propia

Una vez almacenados los datos de las pruebas se listan. El sistema permite realizar un análisis con todas las pruebas registradas o con algunas de ellas, se debe tener en cuenta que como mínimo se deben escoger dos y que se debe ingresar la distancia objetivo para generar un resultado.

Figura 22. Listado de pruebas

Marca	Tiempo	Selección #	Acciones
50.0	41.01	01	Editar Eliminar
50.0	41.02	02	Editar Eliminar
50.0	39.00	03	Editar Eliminar
100.0	19.2	04	Editar Eliminar
150.0	36.71	05	Editar Eliminar

Ingresar la Distancia Objetivo

[Eliminar Marca](#) [Añadir Prueba](#) [Consultar Activo](#) [Mostrar Ayuda](#)

Fuente: elaboración propia

Los resultados que se generan a partir de las pruebas seleccionadas son el ritmo para cada marca registrada, la CSS, el FF y tiempo estimado de acuerdo a las regresiones.

En el software se integraron otros parámetros que permitieran dar una visión más detallada a los nadadores de los diferentes factores que pueden afectar sus entrenamientos, por lo que se incluyeron a partir de las predicciones los cálculos del ritmo estimado y el análisis de ritmos por niveles de Phil Skiba y Bill Sweetenham. Además de la representación a través de gráficas.

Figura 23. Informe de resultados



Fuente: elaboración propia

Por último se realizó un pequeño manual de uso del aplicativo, que se encuentra disponible desde la página principal a través del botón Ayuda que se muestra en la Figura 24.

Figura 24. Instructivo de uso del aplicativo



Fuente: elaboración propia

4.4.3.4. Diseño de vistas

Una vez realizadas las *templates* y generados los resultados se definió las vistas o controladores para cada una de las interfaces. En este punto se hace interesante el uso de la ORM de *Django* pues a través de clases y definiciones se realizaron consultas sencillas, por ejemplo en la

Figura 25 se definen los métodos a partir de clases predefinidas de *Django* para realizar acciones como: listar los registros de las pruebas registradas en la base de datos, el ingreso de nuevos valores, la eliminación y edición de las mismas.

Figura 25. Funciones listar, crear, editar y borrar pruebas

```
class PruebaList(ListView):
    model = Prueba
    template_name = 'pruebas/prueba_list.html'

class PruebaCreate(CreateView):
    model = Prueba
    form_class = PruebaForm
    template_name = 'pruebas/prueba_form.html'
    success_url = reverse_lazy('prueba:listar')

class PruebaUpdate(UpdateView):
    model = Prueba
    form_class = PruebaForm
    template_name = 'pruebas/prueba_form.html'
    success_url = reverse_lazy('prueba:listar')

class PruebaDelete(DeleteView):
    model = Prueba
    template_name = 'pruebas/prueba_delete.html'
    success_url = reverse_lazy('prueba:listar')
```

Fuente: elaboración propia

Los resultados son generados a través de definiciones, que reciben dos variables desde los *templates*, los id de las pruebas seleccionadas en el *checkbox* y que se almacenan en una lista, y la distancia objetivo. A través de la función *filter* de *django*

se identifica con cuales pruebas se desea trabajar y con esas se realiza el cálculo del ritmo:

Figura 26. Captura de datos a través de las vistas

```
def ritmo(request):
    lista = request.POST.getlist('checkbox')
    d = request.POST['distancia_objetivo']
    disobj = int(d)
    # Calculo del ritmo
    a = Prueba.objects.filter(id__in=lista).annotate(a=(F('tiempo')/F('distancia')*100))
```

Fuente: elaboración propia

Como se mencionó en el capítulo anterior la CSS está representada por la pendiente que se calcula a través de la regresión lineal. El algoritmo se implementó de la siguiente manera:

1. Usar la función *aggregate* de *django*, tanto para la distancia como el tiempo, para calcular el promedio de todos los registros existentes, como se muestra en la Figura 27.

Figura 27. Ejemplo de cálculo de la media a partir de la distancia

```
#-----#
# Media de la distancia
mdis = Prueba.objects.filter(id__in=lista).aggregate(dis=(Avg('distancia'))).values()
```

Fuente: elaboración propia

2. Cálculo de la pendiente y el intercepto de las ecuaciones Ecuación 5: Cálculo de la pendiente y Ecuación 6: Cálculo del intercepto de acuerdo a la Figura 28.

Figura 28. Cálculo de la pendiente y el intercepto

```
#-----#
# Calculo de m y b
m = (sdis*stmp - lon*sdt)/(stmp**2 - lon*tmp2)
b = md - m*mt
```

Fuente: elaboración propia

- Una vez obtenidos estos valores se pasa a calcular el tiempo, ritmo estimado y el coeficiente de determinación aplicando la Ecuación 9: Cálculo del tiempo a partir del CSS y ASC como se muestra en la Figura 29.

Figura 29. Tiempo estimado, ritmo estimado, coeficiente de determinación

```
# Calculo del tiempo estimado y el ritmo estimado |
tiemE = (disobj - b) / m
ritmE = tiemE / disobj * 100
#-----#
# Calculo de sigma x
sigmax = math.sqrt(tmp2 / lon - mt**2)
sigmay = math.sqrt(dis2 / lon - md**2)
sigmaxy = (sdt/lon) - (mt*md)
coeficiente = (sigmaxy / (sigmax*sigmay))**2
```

Fuente: elaboración propia

Para el cálculo de la Ecuación 3: Ecuación de resistencia, se realiza el mismo proceso sólo que los datos debes ser transformados a través del logaritmo natural.

Para este caso como los datos requerían un manejo uno a uno se debió realizar la consulta directamente sin hacer uso de la ORM y usar funciones de postgres como AVG, ln o SUM para lograr encontrar los resultados esperados. Como al aplicar este tipo de consultas devuelve un diccionario, se hace necesario convertirlo a una lista para posteriormente recorrerlo:

Figura 30. Cálculos aplicando a la ecuación de resistencia

```
l1 = Brueba.objects.filter(id__in=l1lta).extra(select={
    'tiempo': "(sum(ln (tiempo))/count(''))*((sum(ln (distancia))/sum(ln (tiempo))) *
    (count() ^ (sum(ln (tiempo))/sum(ln (distancia))))))",
    'resistencia': "(sum(ln (distancia))/sum(ln (distancia)))/(count() * (
    (sum(ln (distancia))^ln (distancia))))*(sum(ln (distancia))/count(''))", where=[('tiempo > 0')], values_list('tiempo', flat=True)
```

Fuente: elaboración propia

Se procede a realizar la estimación del ritmo y tiempo a partir del factor fatiga como se muestra en la Figura 31.

Figura 31. Ritmo y tiempo estimado haciendo uso del factor fatiga

```
# Cálculo del tiempo con FF
ritFF = tieFF / disobj * 100
# Tiempo estimado con FF
tieFF = bln * disobj**val
```

Fuente: elaboración propia

Y por último se realizan los cálculos de los valores de los ritmos por nivel de acuerdo a (Martinez A. , 2007)

Figura 32. Cálculo de ritmos por nivel

```
# Calcular la tabla de ritmos estimados por nivel Phil skiba
res = css * 1.07
tem = css * 1.03
umbr = css * 1.01
vo = css * 0.98
cap = css * 0.92

# Calcular la tabla de ritmos estimados por nivel Bill sweetenham
blne = math.exp(valb) # b de ln
test = blne * 200**val
test2 = test / 2
z1a1 = test2 + 20
z1a2 = test2 + 15
z1a3 = test2 + 10
z2= test2 + 7
z3 = test2 + 4
```

Fuente: elaboración propia

Cada resultado obtenido se muestra en la interfaz convertido a horas y minutos, para ello se debió realizar la definición de la función *convertirhora*, la cual recibe el dato como tipo *float* y lo convierte a horas, segundos y minutos.

Figura 33. Definición para convertir a horas, minutos y segundos

```
def convertirhora(ent):
    h = str(int(ent) // 3600)
    minu = str(int(ent) % 3600 // 60)
    r = str(int(ent) % 60)
    hora = h + ':' + minu + ':' + r
    return hora
```

Fuente: elaboración propia

4.4.3.5. Pruebas Funcionales

Las pruebas funcionales se realizaron planteando tres escenarios de prueba con la información de nadadores diferentes que se encuentran registrados en la FECNA.

Del primer nadador se tenían 136 datos, del segundo 28 datos y del tercero 14, como se muestra en la *Figura 34* respectivamente:

Figura 34. Marcas de nadadores 1, 2 y 3

Distancia	Tiempo	Selección	Acciones
50.0	33.55	#1	Editar Eliminar
50.0	34.12	#2	Editar Eliminar
50.0	32.8	#3	Editar Eliminar
50.0	33.19	#4	Editar Eliminar
50.0	32.32	#5	Editar Eliminar
50.0	32.32	#6	Editar Eliminar
50.0	31.69	#7	Editar Eliminar

Distancia	Tiempo	Selección	Acciones
50.0	35.36	#1	Editar Eliminar
50.0	37.11	#2	Editar Eliminar
50.0	37.62	#3	Editar Eliminar
50.0	36.76	#4	Editar Eliminar
50.0	37.4	#5	Editar Eliminar
50.0	36.09	#6	Editar Eliminar
50.0	34.1	#7	Editar Eliminar

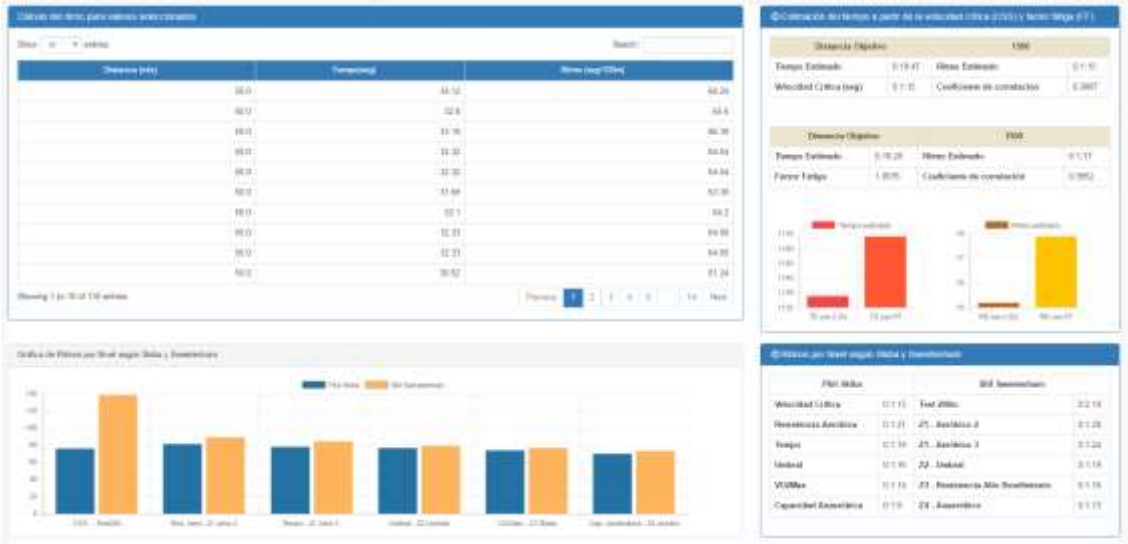
Sistema de evaluación del rendimiento para nadadores colombianos

Distancia	Tiempo	Selección	Acciones
50.0	42.31	#1	Editar Eliminar
50.0	41.07	#2	Editar Eliminar
50.0	39.35	#3	Editar Eliminar
100.0	100.2	#4	Editar Eliminar
100.0	96.75	#5	Editar Eliminar
100.0	82.74	#6	Editar Eliminar
100.0	81.74	#7	Editar Eliminar

Fuente: elaboración propia

Para una distancia objetivo de 1500 mts en el nadador 1 el análisis de resultados se muestra en la Figura 35:

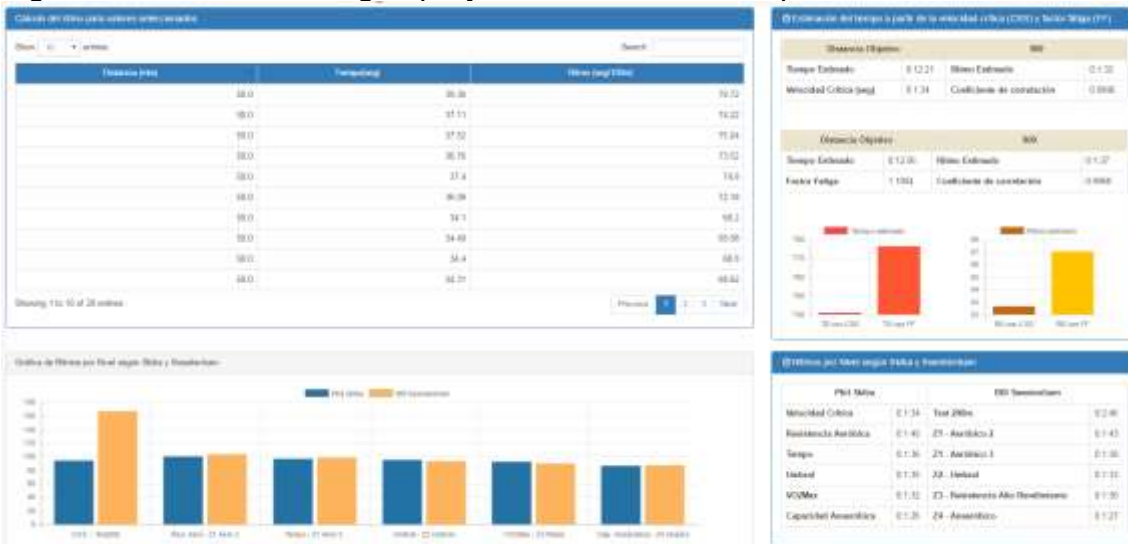
Figura 35. Resultados según proyección del nadador 1 para 1500mts



Fuente: elaboración propia

Análisis del nadador 2 con una distancia objetivo de 800 mts

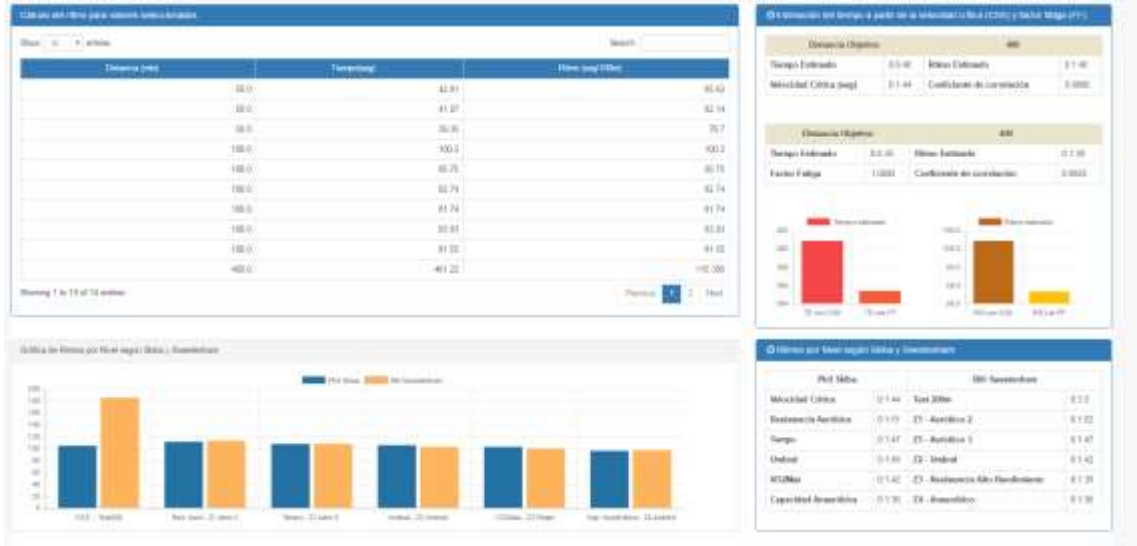
Figura 36. Resultados según proyección de nadador 2 para 800 mts



Fuente: elaboración propia

Análisis nadador 3 con una distancia objetivo de 400 mts como se muestra en Figura 37:

Figura 37. Resultados según proyección de nadador 3 para 400mts



Fuente: elaboración propia

Figura 38. Datos tomados de la FECNA Nadadores 1, 2 y 3

1500m	00:15:14.80	800m	00:09:58.26
Libre/1500m		Libre/800m	
Free		Free	
1500m	00:12:18.33	800m	00:10:15.41
Libre/1500m		Libre/800m	
Free		Free	
1500m	00:12:46.71	800m	00:09:52.96
Libre/1500m		Libre/800m	
Free		Free	
1500m	00:13:59.18	1500m	00:19:54.80
Libre/1500m		Libre/1500m	
Free		Free	
1500m	00:12:51.06		
Libre/1500m			
Free			
400	07:41,2		
400	06:52,7		
400	06:11,9		
400	06:16,5		
400	06:02,3		

Fuente: elaboración propia

Tabla 6 - Comparativo de los resultados registrados por los nadadores y los obtenidos en la predicción

NADADOR	PRUEBA	TIEMPO REGISTRADO	PROMEDIO DE TIEMPO	TIEMPO ESTIMADO CON CSS	TIEMPO ESTIMADO CON FF
Nadador 1:	1500m Libre/1500m Free	00:20:49.82	19:22.39	18:47	19:27
LAURA	1500m Libre/1500m Free	00:19:28.04			
DANIELA	1500m Libre/1500m Free	00:19:17.02			
ABRIL	1500m Libre/1500m Free	00:19:06.23			
LIZARAZO	1500m Libre/1500m Free	00:18:10.85			
Nadador 2:	800m Libre/800m Free	00:09:58.26	09:42.21	12:21	12:56
JOSE	800m Libre/800m Free	00:10:15.41			
ABUCHAIBE MANZANO	800m Libre/800m Free	00:09:52.96			
Nadador 3:	400m Libre/400m Free	00:07:41.22	06:41.54	06:40	06:35
NICOLAS	400m Libre/400m Free	00:06:52.68			
ACEVEDO	400m Libre/400m Free	00:06:11.94			
ACEVEDO	400m Libre/400m Free	00:06:16.45			
	400m Libre/400m Free	00:06:02.32			

Fuente: elaboración propia

5. CONCLUSIONES

El análisis de los datos permitió determinar que el registro de marcas no conserva un comportamiento que se pueda ubicar en una distribución normal, en las gráficas la curvatura de la muestra indica picos y variaciones a medida que se van incrementando la distancia y el tiempo, y en los test de normalidad de Lilliefors se confirmó que los *p-value* son bastante bajos acercándose a cero. Lo anterior además implica que los datos requieren una transformación y el cálculo de más variables que permitan utilizar la media como medida de centralización o modelos lineales simples que faciliten una predicción.

La distribución normal en la muestra se pudo observar cuando se realizó la agrupación de los datos de acuerdo a las distancias registradas de 100, 200, 400, 800 y 1500 mts. Una vez determinada se hizo posible aplicar los modelos de regresión lineal a las ecuaciones planteadas por Wakayoshi y Riegel, con lo que se hizo necesario ampliar la muestra para entrenar el modelo con el propósito de obtener resultados más precisos y acercados a las marcas que puede tener el nadador.

Los modelos de predicción en natación tienden a ser complejos y costosos debido a que el atleta se encuentra sometido a varios factores que influyen en su rendimiento. Por esta razón, en esta investigación se describe una parte complementaria a ese estudio con una propuesta de implementación de un modelo de predicción en natación. El prototipo propuesto se formula a partir de la aplicación de dos modelos de regresión lineal en las ecuaciones de velocidad crítica y de resistencia que considera las posibles influencias del tiempo y la distancia con el fin de analizar y formular estrategias que permitan medir el rendimiento. Además, se formulan otras medidas que son utilizadas actualmente como son los ritmos estimados y por nivel que permiten realizar un análisis complementario para apoyar el sistema de entrenamiento de los nadadores. La efectividad y estrategia de la modelación se muestran con el cálculo del error porcentual medio absoluto o MAPE donde se realiza una estimación separando los datos de entrenamiento con los de predicción y se pudo obtener un error por debajo del 20% de estimación, de manera que los resultados fueron satisfactorios en la predicción y en su representación estadística.

El uso de herramientas como el paquete R y RKWard facilitó el análisis de los datos y la elección de la estrategia de modelamiento, debido a que se pudo aprovechar las características de este tipo de software para manejar grandes volúmenes de datos, tratarlos y realizar sobre ellos operaciones más avanzadas. Además el uso de herramientas software libre durante cada etapa del desarrollo posibilitó su

aprendizaje por toda la documentación disponible además de la inclusión de características en la aplicación como la seguridad, interfaces adaptables y fácil acceso a la información.

6. RECOMENDACIONES Y TRABAJOS FUTUROS

El sistema que se propone en el siguiente proyecto va enfocado a permitirle al nadador realizar un análisis a partir de sus resultados en cuanto a tiempo y distancia, se puede plantear un modelo más cerrado que involucre otras variables, que a su vez puedan generar otros modelos con resultados combinados, estas variables pueden ser: cualitativas, en cuanto a procedencia del nadador, lugar de la prueba, género, edad, estilos de nado y raza, y cuantitativas relacionadas con tiempo de entrenamiento, medidas del nadador, condiciones meteorológicas y estructurales de los sitios donde se realizan las competencias y entrenamientos. Para ello se debe fortalecer las fuentes de almacenamiento de la información recolectada de los nadadores garantizando la integridad de la misma en cuanto a precisión y actualización.

REFERENCIAS Y BIBLIOGRAFÍA

Los modelos matemáticos: un aporte científicotecnológico. (2013). *EFDportes.com Revista Digital*, 1-8.

Metodologías de desarrollo ágil: Lean Development. (20 de 04 de 2017). Obtenido de <http://danielgrifol.es/metodologias-de-desarrollo-agil-lean-development/>

Alvarez, M. A. (19 de 11 de 2013). *Qué es Python*. Recuperado el 08 de 05 de 2016, de <http://www.desarrolloweb.com/articulos/1325.php>

Bisono, T. (23 de 07 de 2014). *Preparación para Competencias: Introducción al Taper*. Obtenido de <http://www.clinicasdenatacionrd.com/el-taper/>

Blundell, J. D. (s.f.). *Numerical Algorithms for Predicting Sports Results*. Recuperado el 06 de 05 de 2016, de <http://www.engineering.leeds.ac.uk/e-engineering/documents/JackBlundell.pdf>

Briega, R. E. (10 de 10 de 2015). *Machine Learning con Python*. Obtenido de <http://relopezbriega.github.io/blog/2015/10/10/machine-learning-con-python/>

Chapman, P. C. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS.

Coulson, M., & CooperyD., J. (30 de 05 de 2011). *alto rendimiento*. Obtenido de <http://altorendimiento.com/test-velocidad-critica-natacion/>

Dobravec, S. (2015). Predicting sports results using latent features: A case study. *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on* (págs. 1267 - 1272). Opatija: IEEE.

Frias, J. F., Martínez, C., Cruz Blandón, M., & Vargas Rojas, L. F. (2014). *Sabio*. Recuperado el 09 de 08 de 2016, de <http://www.sabiofutbol.com/>

- Gaía, A. O. (30 de 10 de 2013). <http://g-se.com/es/entrenamiento-en-natacion/blog/uso-de-la-velocidad-critica-para-el-entrenamiento-de-la-resistencia-aerobica-en-nadadores-jovenes>. Obtenido de Uso de la velocidad crítica para el entrenamiento de la resistencia aeróbica en nadadores jóvenes.: <http://g-se.com/es/entrenamiento-en-natacion/blog/uso-de-la-velocidad-critica-para-el-entrenamiento-de-la-resistencia-aerobica-en-nadadores-jovenes>
- Gómez Valiente, S., & Valdés Corría, G. (01 de 2011). *Federación Nacional de Natación*. Obtenido de <http://www.fecna.com/wp-content/uploads/2011/08/La-T%C3%A9cnica-en-la-Nataci%C3%B3n-Competitiva..pdf>
- Gross, M. (02 de 05 de 2014). *Conozca 3 tipos de investigación: Descriptiva, Exploratoria y Explicativa*. Recuperado el 19 de 04 de 2017, de <http://manuelgross.bligoo.com/conozca-3-tipos-de-investigacion-descriptiva-exploratoria-y-explicativa#.WPd3jGnhCUk>
- Guardiola Jiménez, P. (05 de 06 de 2014). *Universidad de Murcia*. Recuperado el 25 de 07 de 2016, de http://www.um.es/docencia/pguardio/documentos/Tec_logis.pdf
- Guardiola, F. V. (2013). *Sistema de predicción de resultados en eventos deportivos y su aplicación en las apuestas*. Leganés.
- Guazzelli, A. (12 de 07 de 2012). *Predicciones sobre el futuro, parte 2: Técnicas de modelado predictivo*. Recuperado el 12 de 06 de 2016, de <http://www.ibm.com/developerworks/ssa/industry/library/ba-predictive-analytics2/>
- Holovaty, A., & Kaplan-Moss, J. (2015). *El libro de Django 1.8*. Django Software Corporation.
- Ingenio Empresa. (07 de 03 de 2016). *Medición del error en pronósticos de demanda*. Recuperado el 09 de 07 de 2017, de <https://ingenioempresa.com/medicion-error-pronostico/>

- Kyriakides, G., Talattinis, K., & Stephanides, G. (2015). A Hybrid Approach to Predicting Sports Results and an AccuRATE Rating System. *Springer India*.
- Letelier, P., & Penadés, M. C. (2016). Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP). *Ciencia y Técnica Administrativa, ISSN 1666-1680*.
- Mantilla, G. B. (12 de 02 de 2002). *Reglamento de investigación*. Recuperado el 19 de 04 de 2017, de http://www.unab.edu.co/sites/default/files/normatividad_Investigaciones_UNAB/Normatividad/Reglamento%20de%20Investigaciones.pdf
- Martinez, A. (05 de 01 de 2007). *AM triathlon*. Obtenido de <http://www.amtriathlon.com/2007/01/prediccin-del-rendimiento-en-natacin.html#ixzz4dZvqJqxY>
- Martinez, A. E. (21 de 12 de 2012). *ADMARTINEZ*. Obtenido de <http://ironsommelier.blogspot.com.co/2012/12/curva-de-fatiga-en-carrera.html>
- matchstatistics. (06 de 08 de 2009). *matchstatistics.com*. Recuperado el 10 de 05 de 2016, de <http://matchstatistics.soft32.com/>
- Moliner, L. M. (01 de 2001). *LA REGRESION LOGISTICA*. Recuperado el 27 de 07 de 2016, de <http://www.seh-lilha.org/rlogis1.htm>
- Morales, A. T., & Lorenzo Calvo, A. (2012). Análisis de los indicadores de rendimiento en las finales europeas de natación en pruebas cortas y en estilo libre. *Apunts. Educación Física y Deportes* , 97-107 .
- Moreno, S. (21 de 04 de 2012). *Tu mejor plan de entrenamiento*. Obtenido de <http://www.ellocoquecorre.com/2012/04/21/tablas-de-jack-daniels/>
- Morillo, R. A. (01 de 06 de 2015). *PMOinformatica*. Recuperado el 2017 de 05 de 21, de <http://www.pmoinformatica.com/2015/05/historias-de-usuario-ejemplos.html>

- Neuralbelt. (2 de 10 de 2011). *Soccer Match Predictor 1.6.6*. Recuperado el 11 de 05 de 2016, de <http://soccer-match-predictor-1.soft32.com/>
- Ojeda, J. C., & Gómez Fuentes, M. d. (2012). Taxonomía de los modelos y metodologías de. *UDUAL*, 37-47.
- Orellana, L. (2008). *Análisis de regresión*. Obtenido de http://www.dm.uba.ar/materias/estadistica_Q/2011/1/clase%20regresion%20simple.pdf
- Pauly, D. (1983). Algunos métodos simples para la evaluación de recursos pesqueros tropicales. *Algunos métodos simples para la evaluación de recursos pesqueros tropicales*, 243-249.
- RGP, J. (25 de 01 de 2014). *Descubre qué es Django, el framework web de moda*. Recuperado el 10 de 05 de 2016, de <http://computerhoy.com/noticias/internet/descubre-que-es-django-framework-web-moda-8641>
- Riegel, P. (08 de 1977). Athletic Records and Human Endurance, *American Scientist*. *American Scientist*, 285-290.
- Rodríguez Baños, Á., Berral de la Rosa, C. J., Rodríguez Bies, E. C., Lara Padilla, E., & Berral de la Rosa, F. J. (2009). Aplicación de un modelo matemático al taper en jóvenes nadadores. *Archivos de medicina del deporte*, 355-363.
- Rodríguez, D., & Dolado, J. (2007). *Redes Bayesianas en la Ingeniería del Software*. Recuperado el 28 de 07 de 2016, de <http://www.cc.uah.es/drg/b/RodriguezDolado.BBN.2007.pdf>
- Ruiz, A. L., Martínez García, C., Sánchez Pérez, M. J., Sánchez-Cantalejo Castañeda, J., & Sánchez-Cantalejo Ramírez, E. (s.f.). *Aprenda a usar R*. Recuperado el 08 de 05 de 2016, de <http://www.tutorialr.es/es/index.html>
- Sweetenham, B., & Atkinson, J. (2003). *Championship Swim Training*.

Universidad de Cuernavaca. (s.f.). *¡Arma tu ficha bibliográfica!* Recuperado el 19 de 04 de 2017, de http://www.cva.itesm.mx/biblioteca/pagina_con_formato_version_oct/principal%20nueva.html

Universidad de Granada. (s.f.). *Cadenas de Markov*. Recuperado el 28 de 07 de 2016, de http://www.ugr.es/~bioestad/_private/cpfund10.pdf

Uzoma, A. O., & Nwachukwu, E. (2015). A Hybrid Prediction System for American NFL Results. *International Journal of Computer Applications Technology and Research*, 42-47.

Wakayoshi K., I. K. (2012). Determination and validity of critical velocity as an index of swimming performance in the competitive swimmer. *European Journal of Applied Physiology and Occupational Physiology* 64, 153-157.

Zingsem, C. G. (24 de 07 de 2015). *Velocidad crítica como indicador de rendimiento en natación*. Recuperado el 20 de 03 de 2017, de <http://www.mitlabmalaga.com/velocidad-critica-como-indicador-de-rendimiento-en-natacion/>

ANEXOS

Anexo 1 Formato de historias de usuario

Identificador (ID) de la historia	Enunciado de la historia				Criterios de aceptación			
	Rol	Característica / Funcionalidad	Razón / Resultado	Número (#) de escenario	Criterio de aceptación	Contexto	Evento	Resultado / Comportamiento esperado
1	Usuario	Se requiere que el usuario pueda almacenar los resultados de sus pruebas en el sistema	Un formulario donde se permita el ingreso de la distancia y el tiempo	1	Guardar el registro en la base de datos	El usuario ingresa la distancia y el tiempo y da clic en el botón guardar	Guardar	Se guarda el registro en la base de datos
2			Se requiere poder ingresar muchas marcas desde un archivo	2	Guardar varios registros en la base de datos	El usuario carga un archivo .csv y los registros se almacenan en la base de datos	Carga Masiva	Genera un mensaje indicando que el archivo se ha subido correctamente y se visualizan los datos almacenados en la base de datos
3	Usuario	Se requiere listar los registros de la base de datos	Mostrar el resultado de valores que se encuentran registrados en bd	3	Se muestran los datos almacenados en la base de datos	El usuario ingresa al aplicativo y después de guardar un registro este se muestra en pantalla	Listar	Se listan los registros almacenados en la base de datos
4	Usuario	Se necesita generar el resultado a partir de la selección de ciertos registros y el ingreso de la distancia objetivo	Se lista el resultado del tiempo estimado a partir del cálculo de la CSS, FF y otros valores como el ritmo estimado y por nivel	4	Se muestran los valores de tiempo estimado, CSS, FF y ritmos estimados y por nivel	Después de ingresar los registros, el usuario selecciona más de dos y teniendo en cuenta que deben ser con distancias diferentes.	Generar Resultados	Se muestran las estimaciones y gráficas de cada uno de ellos

Anexo 1 (Continuación)

Identificador (ID) de la historia	Enunciado de la historia				Criterios de aceptación			
	Rol	Característica / Funcionalidad	Razón / Resultado	Número (#) de escenario	Criterio de aceptación	Contexto	Evento	Resultado / Comportamiento esperado
5	Usuario	Se requiere eliminar registros de la base de datos	Eliminar un registro de la base de datos	5	El registro no aparece en listado ni en la base de datos	El usuario selecciona el registro que desea eliminar, da clic en el botón eliminar y confirma que está seguro	Eliminar	Se elimina el registro de la base de datos
6	Usuario	Se requiere editar registros de la base de datos	Se muestra los cálculos del análisis	6	Se permite eliminar un registro ya sea en la distancia o el tiempo	El usuario selecciona el registro que desea editar, da clic en el botón editar y modifica algún campo. Por último da clic en guardar	Editar	El registro se editó
7	Usuario	Se necesita un manual de ayuda del aplicativo	Se muestra los cálculos del análisis	7	Se muestra un área de ayuda para uso del aplicativo	El usuario accede a la ayuda y se muestran capturas y descripciones de cada una	Ayuda	Se muestra los pasos y capturas para facilitar el uso del aplicativo