



Ingeniería en Informática

*“Sistema de toma de decisión para las carreras de la
universidad Columbia utilizando Machine Learning y Minería
de datos”*

Alumno:

Carlos Enrique Apuril Balbuena

Tutor: Msc. Roberto Sánchez

Línea de Investigación:

Minería de datos

SAN LORENZO – PARAGUAY

2022



ÍNDICE

Contenido.....	V
Lista de tablas	VI
Lista de figuras	VI
Lista de abreviaturas	VII
Resumen	8
Introducción.....	9
Planteamiento y descripción del problema	10
Problema	10
Preguntas de la investigación.....	10
Específicas	10
Objetivo General.....	11
Objetivos Específicos	11
Bases teóricas.....	12
Análisis predictivo:	15
Machine Learning	16
Data Mining	17
Algoritmo k- Nearest Neighbors	21
Arboles de decisión.....	23
Naive Bayes	25
Regresión Lineal.....	27
Weka.....	28
Como utilizar Api de Weka Java	31
¿Por qué una aplicación web?.....	35
Antecedentes	35
Especificación de la solución.....	36
Alcance	36
Especificaciones y requerimientos	37
Arquitectura	39
Estructura de base de datos	40
Obtención de datos del dataset	41
Metodología	43
Tipos de investigación	43
Aporte de la solución	44
Demostración del circuito completo del sistema y análisis de los datos.	44
Conclusión	52
Recomendaciones	52
Referencias	53



RESUMEN

La elección de una carrera universitaria es un proceso personal que no debe tomarse con apresuramiento. Algunos errores que se cometen es que muchos estudiantes escogen una carrera universitaria sin tener muy claro de qué tratan esos estudios, y sin profundizar en los detalles para ver si estos se ajustan a la personalidad y gusto del estudiante. Algunos creen que conocen bien los contenidos de la carrera universitaria que se disponen a cursar, y luego se encuentran algo muy diferente de lo que pensaban. Otros deambulan de carrera en carrera sin hallar su verdadera pasión. Es por todo esto que el presente trabajo consiste en ayudar a la toma de decisión para la elección de una carrera universitaria, utilizando herramientas de data mining y machine learnig. La solución que se plantea es la implementación de una plataforma web que tenga una interfaz amigable e interactiva, la cual a través una serie de preguntas al usuario permita dar resultados predictivos mediante distintos algoritmos. Los resultados obtenidos son empleados en el sitio web para mostrar al usuario la predicción de la carrera universitaria y además le permite ver detalles relevantes adicionales como proyecciones futuras de porcentaje de posibilidad de encontrar empleo y promedio de salario, entre otros datos de la carrera universitaria.