UNIVERSIDAD PERUANA UNIÓN

FACULTAD DE INGENIERÍA Y ARQUITECTURA

Escuela Profesional de Ingeniería de Sistemas

Enfoque plano-jerárquico basado en modelo de aprendizaje automático para la clasificación de productos de e-commerce

Tesis para obtener el Título Profesional de Ingeniero de Sistemas

Autor:

Harold Enrique Cotacallapa Mamani

Asesor:

Mg. Nemias Saboya Rios

Lima, noviembre de 2023

DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo Nemias Saboya Rios, docente de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **"ENFOQUE PLANO-JERÁRQUICO BASADO EN MODELO DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DE PRODUCTOS DE E-COMMERCE"** del autor Harold Enrique Cotacallapa Mamani tiene un índice de similitud de 4% verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima, a los 23 días del mes de noviembre del año 2023.

Mg. Nemias Saboya

Firmado digitalmente por Mg. Nemias Saboya
DN: C=PE, OU=FIA, O=UPeU, CN=Mg. Nemias Saboya, E=saboya@upeu.edu.pe
Razón: Soy el autor de este documento
Ubicación:
Fecha: 2023.11.30 14:48:29-05'00'
Foxit PDF Reader Versión: 12.1.0

Mg. Nemias Saboya Rios

ACTA DE SUSTENTACIÓN DE TESIS

En Lima, Ñaña, Villa Unión, a los **15** día(s) del mes de **noviembre** del año 2023 siendo **las 14:00 horas**, se reunieron en modalidad virtual u online sincrónica, bajo la dirección del Señor Presidente del jurado: **Dra. Erika Inés Acuña Salinas**, el secretario: **Ph.D. Javier Linkolk Lopez Gonzales**, y los demás miembros**: Mg. Danny Levano Rodríguez** y **el MSc. Fredy Abel Huanca Torres**, y el asesor, **Mg. Nemias Saboya Rios**, con el propósito de administrar el acto académico de sustentación de la tesis titulada: " Enfoque plano-jerárquico basado en modelo de aprendizaje automático para la clasificación de productos de e-commerce"

de el(los)/la(las) bachiller/es: a) **HAROLD ENRIQUE COTACALLAPA MAMANI**

…… b) **……**

conducente a la obtención del título profesional de **INGENIERO DE SISTEMAS**
(Nombre del Título profesional)

con mención en…….....

El Presidente inició el acto académico de sustentación invitando al (los)/a(la)(las) candidato(a)/s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por el(los)/la(las) candidato(a)/s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato (a): ..... **HAROLD ENRIQUE COTACALLAPA MAMANI** ……

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
APROBADO	20	A	EXCELENCIA	EXCELENCIA

Candidato (b): ..... **……** ……

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó al(los)/a(la)(las) candidato(a)/s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

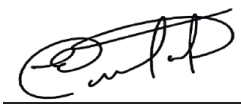
Presidente
Dra. Erika Inés Acuña Salinas

Secretario
PhD. Javier Linkolk Lopez Gonzales

Asesor
Mg. Nemias Saboya Rios

Miembro
Mg. Danny Lévano Rodríguez

Miembro
MSc. Fredy Abel Huanca Torres



Candidato/a (a)
Harold Enrique Cotacallapa Mamani

Candidato/a (b)

ÍNDICE

Enfoque plano-jerárquico basado en modelo de aprendizaje automático para la clasificación de productos de e-commerce

A flat-hierarchical approach based on machine learning model for e-commerce product classification

RESUMEN

En el ámbito del comercio electrónico, optimizar el proceso de clasificación de productos adquiere una importancia crucial debido a su influencia directa en la eficiencia operativa y, por ende, en la rentabilidad. Aunque se han dedicado esfuerzos académicos considerables para abordar este problema, persisten lagunas en la literatura existente. En tal sentido, este artículo presenta una solución para la clasificación jerárquica de productos de comercio electrónico usando un conjunto de datos de 4 niveles de profundidad, obtenidos de una destacada plataforma de comercio electrónico en América Latina. Nuestra propuesta consiste en un modelo de aprendizaje automático que integra enfoques tanto el clasificación plana como local (jerárquica) para mejorar la eficacia individual de cada uno. En busca de este objetivo, se llevó a cabo un análisis comparativo de siete algoritmos de aprendizaje automático: Multinomial Naive Bayes Multinomial, Linear Support Vector Classifier, Multinomial Logistic Regression, Random Forest, XGBoost, FastText y Voting Ensemble. Los tres primeros se utilizaron para el modelo que emplea el enfoque local, mientras que el modelo que usa el enfoque plano es el Voting Ensemble compuesto por los 3 primeros algoritmos mencionados anteriormente. Los resultados demostraron que este enfoque plano-jerárquico superó al mejor modelo de enfoque plano en un 0.15% y al mejor modelo de enfoque local (Clasificador Local por Nivel) en un 4.88%, medido por el puntaje F1 ponderado. Además, se pone a disposición un nuevo conjunto de datos en español con más de un millón de productos de comercio electrónico. Finalmente, se discuten las mejores técnicas de preprocesamiento para este conjunto de datos, junto con las limitaciones inherentes al estudio y las posibles direcciones para futuras investigaciones en esta área.

ABSTRACT

Within the e-commerce sphere, optimizing the product classification process assumes pivotal importance, owing to its direct influence on operational efficiency and, by extension, profitability. While extensive scholarly efforts have addressed this issue, persistent gaps remain within the existing literature. Therefore, this paper introduces a solution for hierarchical classification using a 4-level electronic product dataset obtained from a renowned e-commerce platform in Latin America. Our proposal consists of a Machine Learning model that integrates both flat and local (hierarchical) classification approaches to enhance each individual's efficacy. In pursuit of this goal, a comparative analysis of seven machine learning algorithms, including Multinomial Naive Bayes, Linear Support Vector Classifier, Multinomial Logistic Regression, Random Forest, XGBoost, FastText, and Voting Ensemble, was conducted. The first three were used for the model employing the local approach, while all seven were used for the model with the flat approach. The results demonstrated that this flat-hierarchical approach outperformed the best flat approach model by 0.15\% and the best local approach model (Local Classifier per Level) by 4.88\%, as measured by the weighted F1-score. Additionally, a new dataset in Spanish with over one million e-commerce products is made available. Finally, the best preprocessing techniques for this dataset are discussed, along with the study's inherent limitations and future research directions in this field.