

DATA CLASSIFICATION WITH MACHINE LEARNING ALGORITHM IMPLEMENTATION ON DISTANCE LEARNING EXPERIENCES

**A Thesis Presented to the Faculty of the
College of Communication and Information Technology
President Ramon Magsaysay State University
Castillejos, Zambales**

**In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science**

By:

**Alcea Katigbak
Roderick G. Monsalve Jr.
Mutya C. Ronquillo**

June 2023

Republic of the Philippines
PRESIDENT RAMON MAGSAYSAY STATE UNIVERSITY
(Formerly Ramon Magsaysay Technological University)
 Castillejos Campus
 Castillejos, Zambales



APPROVAL SHEET

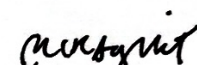
The thesis project entitled **“Data Classification with Machine Learning Algorithm Implementation on Distance Learning Experiences”** was prepared and submitted by **Alcea Katigbak, Roderick G. Monsalve Jr., and Mutya C. Ronquillo** in partial fulfillment of the course requirements for the degree of **Bachelor of Science in Computer Science** has been examined and recommended for the oral examination.


MICHAEL G. ALBINO, MIT
 Thesis Adviser

Approved by the Panel of Examiners
 with a rating of 94 %

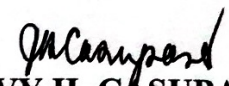

IRATUS GLENN A. CRUZ, LPT, MSCS
 Chair


MICHAEL N. FARIN, MSCS
 Member


MARIE CELIA R. AGLIBOT, MSCS
 Member

Accepted and approved as a requirement for the degree of
BACHELOR OF SCIENCE IN COMPUTER SCIENCE.

June 2023
 Date


IVY H. CASUPANAN, EdD
 Campus Director

ABSTRACT

Sentiment Analysis has become a great tool for identifying the sentiments of social media users not only in product branding but most especially in determining sentiments in education. The researchers focused on Data Classification techniques (DCL) using Natural Language Processing (NLP) techniques to identify social media sentiments on the topic of "Distance Learning". The sentiment analysis determined the positive, negative, and neutral statements of the social media users using the Twitter Application programming interface (API).

The researchers tested the trained sentiment model through several machine learning algorithms but not limited to K-Nearest Neighbors (K-NN), Tree (T), and Naïve Bayes (NB) which yielded precision and recall values of 0.996 (K-NN), and 0.999 (Naïve Bayes and Tree). It can be noted that the trained model has a positive result to the selected three (3) algorithms in terms of its efficiency. On the other hand, the confusion matrix shows an accurate prediction of the trained models as reflected in the result of the testing using the confusion matrix.

The researchers, therefore, concluded that the result of the model training has a positive correlation to the extracted keywords on the datasets. Furthermore, the evaluation of the Liu & Hu and VADER sentiment analysis helped the researchers in the prediction of the sentiments. It is therefore recommended that sentiment analysis can be used to identify the sentiments of social media users and can help in future planning in distance learning education.

Keywords: *Data Classification technique; Natural Language Processing; Sentiment Analysis*