INFORMATICS INSTITUTE OF
TECHNOLOGY
*IN COLLABORATION WITH*

ROBERT GORDEN UNIVERSITY - UK

(MSc) Big Data Analytics
Final Year Project 2023

# Personality Type Identification System using Unstructured text in English

# based on Natural Language Processing and Machine Learning

A Dissertation By

Pamodya Jayangani De Seram

IIT ID – 20211035

RGU ID -2119027

Supervised By
Mr. Pumudu Fernando

## Abstract

Perosnality type identification is beneficial to understand the associates. Especially to understand life partner in order to have a successful marriage life, select the most suitable candidates for a company, and understand and explore the self capacities are some of them. Personality is combination of a person's behavior, feelings, motivation, and thought patterns. Those characteristics take years to understand and identify in a person's personality. Personality type identification system was proposed to speed up the process.

In the study, the author identified that existing systems have a gap in identifying personality using unstructured text. To provide speedy and accurate information, the author selected to use Natural Language Processing and Machine Learning techniques.

Therefore, the author used the Decission Tree algorithm, the K-Neigbour Algorithm, Support Vector Machine, Naive Bayes algorithm, Logistic Regression algorithm, Random Forest algorithm, XGBoost model, and the LightGBM model. Considering the data set analysis, algorithms' accuracy and evaluation metrics, the author developed Voting classifier ensemble model. The author balanced the dataset to improve the accuracy, and achieved 64.48%. The evaluation process carried out with different criteria and groups. The overall evaluator feedback was positive.

*Keywords : Personality Type Identification System, Myers Briggs Type Indicator, Natural Language Process, Machine Learning, Classification, Ensemble model*