

ABSTRACT

MODELING ALZHEIMER'S DISEASE (AD) RISK PREDICTION USING THE RANDOM FOREST ALGORITHM

By:

ANISA NOVIANA

E-mail: anisanoviana211@gmail.com

Alzheimer's Disease (AD) is a leading cause of dementia and ranks as the seventh highest cause of mortality worldwide. It is characterized by progressive deterioration in memory, cognitive abilities, and daily functioning due to irreversible damage to the brain's nervous system. This study aims to evaluate the effectiveness of the Random Forest algorithm in predicting the risk of Alzheimer's Disease by integrating the Pearson Correlation feature selection method to identify the most influential variables. The dataset used in this research was obtained from Kaggle, comprising 2,149 records with 32 features and one target variable. The analysis was conducted using Google Colaboratory within a binary classification framework. The results demonstrate that the Random Forest model achieved high predictive performance, with an accuracy of 95.58%, precision of 96.53%, recall of 90.89%, and an F1-score of 93.60%, despite the presence of class imbalance in the dataset. These findings confirm that the Random Forest algorithm is an effective and reliable approach for predicting Alzheimer's Disease risk, offering a promising decision-support tool for early detection and diagnostic assistance in clinical settings.

Keywords: Alzheimer's Disease, Random Forest, Machine Learning, Risk Prediction, CRISP-DM

