



**PROJECT REPORT**  
**PREDICTING FLIGHT DELAY USING RANDOM**  
**FOREST ALGORITHM, XGBOOST ALGORITHM, AND**  
**STACKING ENSEMBLE METHOD**

**FERREY ADINARTA**  
**21.K1.0007**

**Faculty of Computer Science**  
**Soegijapranata Catholic University**  
**2024**

## ABSTRACT

*Flight delays are problematic for both passengers and airlines. With the increasing amount of flight traffic volume, time punctuality is important since it significantly influences passengers' satisfaction and airline companies' financial performance. Many studies have been conducted to predict these delays by using machine learning algorithms. In some research, it was found that combining more than one machine learning algorithm can improve the prediction results. Therefore, in this research, a comparison of machine learning ensemble methods like bagging, boosting, and stacking to predict flight delays is compared. The objective of this research is to find the best-performing ensemble method for flight delay prediction. A dataset from Kaggle named 'Flight Status Prediction' is used as the dataset for this research. Then, the dataset is cleaned and modified using the preprocessing steps. After that, the dataset is fitted to each ensemble model using the Random Forest algorithm as the bagging method, the Extreme Gradient Boosting (XGBoost) algorithm as the boosting method, and combining both algorithms using the stacking method with Random Forest as the first learner, and the results are evaluated based on the accuracy, recall, and precision values. The results are gotten from two different dimensional reduction methods, which are feature selection and principal component analysis (PCA). The results obtained from this study are that the XGBoost model performs best on predicting flight delays with a mean average accuracy of above 95% in both dimensionality reduction methods, while the Stacking Ensemble method performs the worst with a mean accuracy of less than 92% in both dimensionality reduction methods.*

*Keyword: Flight delay prediction, Ensemble method comparison, Random Forest, Extreme Gradient Boosting (XGBoost), Stacking Ensemble Method, Principal Component Analysis, Feature Selection*