**NATIONAL UNIVERSITY HO CHI MINH CITY**

**UNIVERSITY OF INFORMATION TECHNOLOGY**

**FACULTY OF INFORMATION SYSTEM**

-----🙞🙜🕮🙞🙜-----

A blue logo with a black background

Description automatically generated

**TOPIC: FLIGHT PREDICTION**

**BIG DATA**

Instructor: **MSc. Nguyen Ho Duy Tri**

Class: **IS405.O22.HTCL**

**1. Summary of topic content**

This report investigates the use of linear regression for predicting flight delays, utilizing a dataset obtained from Kaggle. The study focuses on building a predictive model to estimate delays based on various flight-related features such as departure and arrival times, flight duration, airline information, and historical delays. The process involves data preprocessing steps including cleaning, encoding categorical variables, and normalizing the data. Significant features are selected to improve the model's accuracy. The linear regression model is then trained and evaluated by using metrics. The results of the predictions are presented in detail in the report, showcasing the model's ability to predict flight delays, also discusses the model's performance, potential improvements, and implications for real-world applications in the aviation industry.