# Predictive Analytics for ADF Personnel Separation

| | |
|---|---|
| **Title** | Predictive Analytics for ADF Personnel Separation |
| **Description** | We aim to develop an ADF Personnel Separation Prediction model using a rich dataset of de-identified personnel data, encompassing various parameters such as age, rank, marital status, years of service, and time spent at rank. The objective is to apply advanced Machine Learning algorithms in conjunction with Operational Research techniques, particularly proportional hazard models like the Cox model. The project will focus on predicting which personnel are likely to separate within the next 6-12 months, identifying key factors driving separation, and determining additional data that can enhance prediction accuracy. This model will provide valuable insights for strategic human resource planning and retention strategies in the military context. |
| **Expected outcomes for the project upon completion** | Objectives:<br>1. Development of a Robust Prediction Model: To create a sophisticated Machine Learning model that accurately predicts which ADF personnel are likely to separate within the next 6-12 months, leveraging the diverse parameters available in the de-identified personnel data.<br>2. Identification of Key Separation Factors: To analyse and identify the most significant factors contributing to personnel separation, providing insights into aspects like age, rank, marital status, and service duration that most influence decision-making.<br>3. Enhancement of Prediction Accuracy through Data Exploration: To explore and incorporate additional relevant data sources (such as performance data) to improve the accuracy and comprehensiveness of the prediction model, thus aiding in more informed and strategic decision-making.<br>Project Deliverables:<br>1. Predictive Separation Model: A fully developed and tested machine learning model capable of predicting ADF personnel separation. Success is measured by the model's accuracy in forecasting separations 6-12 months in advance, evaluated against a validation dataset.<br>2. Factor Analysis Report: A comprehensive report detailing the key factors influencing personnel separation, identified through data analysis. Success is indicated by the depth and clarity of insights provided about the most impactful variables.<br>3. Data Enhancement Strategy: A strategic plan outlining potential additional data sources and methods for incorporating them to improve the model's accuracy. Success looks like a clearly defined roadmap for future data integration, with identified sources and implementation steps. |

| | |
|---|---|
| **Preparation required by students prior to initial meeting (if any)** | Prior to our initial meeting, students are expected to familiarize themselves with the concepts of employee attrition prediction and common machine learning models. This includes understanding the fundamentals and applications of Logistic Regression, Random Forest, and Decision Trees. This preparation will enable students to have a foundational understanding of the techniques and approaches that will be relevant for developing the predictive model for ADF personnel separation. |
| **Additional Notes by Host Organisation** | Description of the Data students will be working with:<br>The data set students will be working with comprises close to 50,000 records, all de-identified, stored in an Excel sheet. This extensive dataset provides a rich foundation for analysis, ensuring comprehensive coverage and diversity in the variables related to ADF personnel. The de-identification of data ensures ethical compliance and privacy protection, while the Excel format offers ease of access and manipulation for various analytical techniques. |
| **Course Name** | Master of Data Science |
| **What datasets are available and in what format will they be provided?** | The dataset for the project will be provided in a structured format, specifically as an Excel file (.xlsx). This format is ideal for handling large datasets like the one with close to 50,000 records, as it allows for efficient organization and easy manipulation of data. Excel's compatibility with various data processing and machine learning tools ensures a seamless workflow for the students. The data being de-identified ensures ethical use and privacy compliance. Excel's widespread use and familiarity among data science students will facilitate an efficient start to the project. |
| **Industry Contact** | Mr Rick Doblanovic |
| **Host Organisation Contact Email** | rick.doblanovic@c4engineering.com.au |
| **Host Organisation Contact Phone** | 0406974007 |
| **Host Organisation Contact Position** | Managing Director |
| **Host Organisation Name** | C4 Engineering Pty Ltd |
| **Industry Address** | 406/434 St Kilda Rd, Melbourne , Victoria 3004 Australia |
| **Students** | Yufeng Liu, Junheng YU, Pengjiabei TANG, Cheng Qian |
| **Supervisors** | Wanting He |