Project and Data Management (PDM) Plan

1. Project Overview

Project Title:

Predicting Employee Attrition and Layoff Trends Using Machine Learning

Summary of Project Topic and Background:

Employee attrition has become a significant challenge for organizations worldwide. Understanding why employees leave, whether due to voluntary resignation or layoffs, is crucial for businesses to improve retention strategies and workforce planning. Recent global layoffs due to economic downturns and automation further emphasize the need for data-driven decision-making. This project aims to leverage machine learning techniques to predict attrition trends based on factors such as job satisfaction, salary, work-life balance, and company policies.

Research Question:

What are the key factors influencing employee attrition and layoffs, and how accurately can machine learning predict attrition trends across different industries?

Project Objectives:

1. **Data Collection & Processing:** Gather real-time employee attrition data from various sources.

2. **Feature Engineering:** Identify and preprocess key factors influencing attrition.

3. **Machine Learning Implementation:** Develop predictive models using classification, clustering, and regression techniques.

4. **Performance Evaluation:** Compare model accuracy and insights for practical HR applications.

5. **Ethical Compliance:** Ensure GDPR and organizational policies are followed in data handling.

Reference List:

1. Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). **"One hundred years of employee turnover theory and research."** Journal of Applied Psychology. Available Online

2. Gupta, M., & Gupta, D. (2021). **"Predicting Employee Attrition using Machine Learning Algorithms."** International Journal of Engineering Research and Applications. Available Online

3. Saks, A. M. (2006). **"Antecedents and consequences of employee engagement."** Journal of Managerial Psychology. Available Online

2. Project Plan: Task List and Timeline

Task List with Dates:

Task	Description	Date
Project Proposal	Define research question, scope, and objectives.	Week 1
Literature Review	Gather relevant studies and refine methodology.	Week 2-3
Dataset Collection	Collect data from online sources and HR databases.	Week 4
Data Preprocessing	Clean, normalize, and transform data for ML models.	Week 5-6
Feature Selection & Engineering	Identify key attributes for predictive modeling.	Week 7
Model Selection & Implementation	Apply classification, clustering, and regression models.	Week 8-9
Evaluation & Optimization	Assess model accuracy and improve performance.	Week 10
Ethical Review & Compliance	Ensure data privacy and ethical usage.	Week 11
Final Report & Presentation	Summarize findings and prepare for submission.	Week 12

Project Timeline (Gantt Chart)

Project Timeline - Gantt Chart
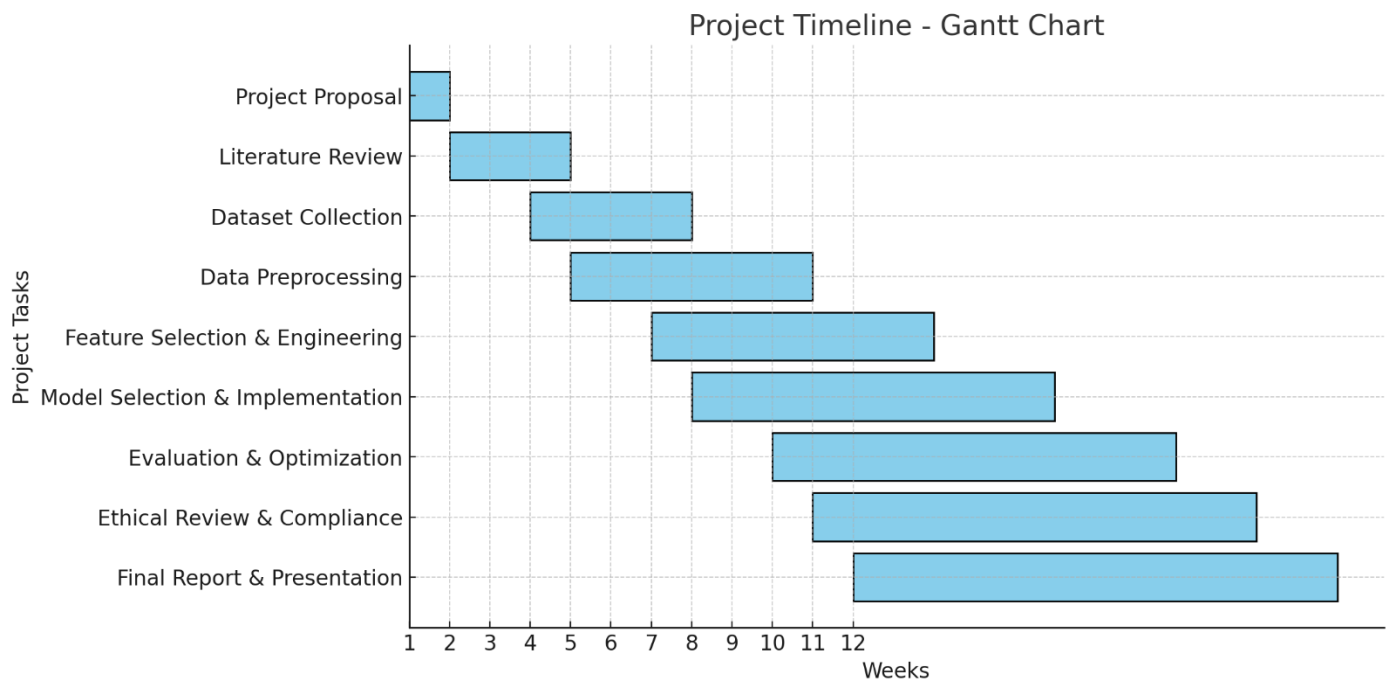
3. Data Management Plan

Overview of Dataset:

The dataset will contain employee demographic details, job satisfaction scores, work experience, salary details, layoff records, and exit interview feedback. Data will be sourced from publicly available datasets, HR repositories, and surveys.

Data Collection Sources:

- Kaggle: Employee attrition datasets
- Open-source HR databases
- Government labor reports
- LinkedIn/Glassdoor job market analytics

Metadata:

- **Format:** CSV, JSON
- **Size:** Estimated ~500MB (structured data)
- **Records:** ~100,000 employee records
- **Attributes:** Employee ID, Age, Department, Salary, Work Experience, Job Satisfaction, Layoff Reason, etc.

Version Control & GitHub Repository:

- **GitHub Repository:** https://github.com/MRashi1/Final-Project
- **Commit Frequency:** Weekly updates with version tagging
- **File Naming Convention:** CSV, ipynb
- **ReadMe File:** Overview of dataset, preprocessing steps, model details, and instructions for reproducibility.

Security and Storage:

- **Backup Plan:** Weekly backup on OneDrive and external storage
- **Data Sharing:** Limited access for supervisor and markers
- **Compliance:** Follows GDPR and UH ethical policies

Ethical Considerations:

1. **GDPR Compliance:** The dataset will be anonymized to ensure no personally identifiable information (PII) is stored. All data handling procedures will align with GDPR standards to protect employee privacy.

2. **UH Ethical Policies:** The project follows the ethical guidelines set by the University of Hertfordshire, ensuring that all research activities comply with institutional ethical standards.

3. **Permission to Use Data:** The dataset used will be obtained from publicly available sources or with explicit consent from data providers. Any proprietary datasets will only be used with legal and documented permission.

4. **Data Integrity and Transparency:** The sources of data will be verified to ensure reliability, and data collection methods will be clearly documented. Any preprocessing steps will be recorded for full transparency.

5. **Bias Mitigation:** Machine learning models will be evaluated for bias to ensure that no demographic group is unfairly impacted. Steps such as fairness metrics and bias correction techniques will be incorporated.

6. **Security Measures:** The dataset will be stored securely on encrypted storage solutions with access control to prevent unauthorized modifications or data breaches.

7. **Ethical Data Usage:** The results of this study will not be used to discriminate against employees but rather to provide insights that can help organizations develop fair and inclusive policies to improve employee retention and job satisfaction.

Student Name: Rashi Meda

Student ID: 23005003

Supervisor Name: William Alston