Summary Report on Employee Attrition Prediction

Introduction:

Employee attrition, the phenomenon of employees leaving an organisation, is a critical concern for businesses as it can impact productivity, morale, and overall organisational performance. In this analysis, we utilised machine learning techniques to predict employee attrition using the **IBM HR Analytics Employee Attrition dataset.**

Data Preprocessing:

The dataset was preprocessed to handle missing values (none found) and **encode categorical variables using Label Encoding**. Numerical **features were scaled using StandardScaler** to ensure all features contribute equally to the model.

Model Selection and Evaluation:

Three classification algorithms were employed: **Logistic Regression, Random Forest, and Support Vector Machine (SVM)**. Each model's performance was evaluated using precision, recall, and F1-score metrics for both classes (0: Not Attrition, 1: Attrition).

Logistic Regression: Achieved an **accuracy of 89%**. While precision and recall for class 0 (Not Attrition) were high, recall for class 1 (Attrition) was relatively low at 21%, indicating the model's struggle to correctly identify instances of attrition.

Random Forest: Achieved an **accuracy of 88%**. Similar to Logistic Regression, it showed high precision and recall for class 0, but recall for class 1 was only 8%, indicating poor performance in identifying attrition cases.

Support Vector Machine (SVM): Achieved an **accuracy of 87%**. Again, while precision and recall for class 0 were high, recall for class 1 was 0%, indicating an inability to identify instances of attrition.

Hyperparameter Optimization:

GridSearchCV was employed to optimise hyperparameters for each model to potentially improve performance.

Optimised Random Forest: Achieved an **accuracy of 88%**. The optimization resulted in slight improvements in precision, recall, and F1-score for class 1 (Attrition), although still relatively low.
Optimised Logistic Regression: Achieved an **accuracy of 89%**. The optimization led to improvements in precision, recall, and F1-score for class 1 (Attrition), indicating better performance in identifying attrition cases compared to the base model.

Optimised SVM: Achieved an **accuracy of 87%**. However, similar to the base model, it failed to identify instances of attrition, with recall for class 1 remaining at 0%.

Insights:

Logistic Regression outperformed the other models in identifying instances of attrition after hyperparameter optimization.
Class imbalance (more instances of non-attrition than attrition) may have affected model performance, especially in correctly identifying instances of attrition.
Further feature engineering and selection could potentially improve model performance by capturing more relevant patterns related to attrition.
Conclusion:

In conclusion, while machine learning models showed promising results in predicting employee attrition, **further refinement and feature engineering are necessary to enhance model performance**, particularly in correctly identifying instances of attrition. Additionally, **addressing class imbalance and exploring alternative modelling techniques could provide further insights** into predicting and potentially mitigating employee attrition effectively.