IIT Bombay

22M1855

Home Logout IITB Website

Roll no 22M1855
Name Aditya Mahendra Satpute
Department Metallurgical Engineering and Materials Science
Program M.Tech (Process Engineering)

Payment | Performance Summary | Graduation Requirements | Personal Information | Forms/Requests

Academic Performance Summary

Year	Sem	SPI	CPI	Sem Credits Used for SPI	Completed Semester Credits	Cumulative Credits Used for CPI	Completed Cumulative Credits
2022	Spring	8.82	8.98	22.0	22.0	57.0	57.0
2022	Autumn	9.09	9.09	35.0	35.0	35.0	35.0

Semester-wise Details

*This registration is subject to approval(s) from faculty advisor/Course Instructor/Academic office.

Year/Semester: 2023-24/Autumn

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CS 725	Foundations of Machine Learning	6.0	Core course	Not allotted	A

Year/Semester: 2023-24/Project

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
MM 797	I Stage Project	50.0	Core course	Not allotted	C

Year/Semester: 2022-23/Spring

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
MM 656	Simulation and Optimisation	6.0	Department elective	AA	C
MM 674	Materials & Processes for Semiconductor Devices	6.0	Department elective	BB	C
MM 677	Diffusion and Kinetics	6.0	Core course	AB	C

MM 694	Seminar	4.0	Core course	BB	C
MM 749	Statistics and Probability for Materials Engineers	6.0	Additional Learning	CD	C
MM 899	Communication Skills	6.0	Core course	PP	N
TA 101	Teaching Assistant Skill Enhancement & Training (TASET)	0.0	Core course	PP	N

Year/Semester: 2022-23/Autumn

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
GC 101	Gender in the workplace	0.0	Core course	PP	N
MM 641	Numerical Methods in Materials Processing	6.0	Department elective	AB	C
MM 651	Thermodynamics of Materials	6.0	Core course	AA	C
MM 659	Transport Phenomena	6.0	Core course	BB	C
MM 680	Welding Science and Technology	6.0	Department elective	AB	C
MM 731	Experiments in Advanced Materials Processing (Lab)	5.0	Core course	AB	C
MM 732	Structural Characterization of Materials First Half	3.0	Core course	AB	C
MM 733	Mechanical Characterization of Materials Second Half	3.0	Core course	AA	C

Report Problem

Name: Aditya Mahendra Satpute

MM749 – Probability and Statistics for Materials Engineers

Roll No. 22M1855

Project: BANK CHURN PREDICTION

Introduction:

The following report outlines the methodology, analysis, and results of a project aimed at predicting customer churn in a banking dataset. The project employed various techniques, including data analysis, data preprocessing, dimensionality reduction, and machine learning algorithms. The goal was to develop an accurate churn prediction model and enhance understanding through visualization.

Methods:

1. **Data Analysis:** Univariate and bivariate analyses were conducted to understand the data distribution and relationships between variables. Kurtosis of curves was examined to assess the data's tail behavior and potential outliers.

2. **Data Transformation:** Appropriate transformations were applied to address data skewness and improve model performance. These transformations played a crucial role in preparing the data for modeling.

3. **Addressing Imbalance:** To tackle class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was employed. This technique generated synthetic samples to balance the distribution of the target variable, thereby enhancing the model's ability to learn from both classes.

4. **Categorical Feature Encoding:** One-Hot Encoding was utilized to convert categorical features into a numerical format suitable for machine learning algorithms. This approach ensured that categorical variables were properly incorporated into the models.

5. **Dimensionality Reduction and Visualization:** Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset while retaining relevant information. This led to a reduced feature set with 72% variance explained. Visualization techniques were employed to present the reduced-dimensional data effectively.

6. **Modeling:** Three different machine learning algorithms were implemented to predict churn: Random Forest (RF), Adaboost, and Support Vector Machine (SVM). The F1 score was chosen as the evaluation metric due to the class imbalance. These models were selected for their ability to handle complex relationships and provide accurate predictions.

Results:

The project achieved promising results in predicting customer churn:

- Random Forest: F1 score of 0.91
- Adaboost: F1 score of 0.89
- SVM: F1 score of 0.89

The confusion matrices for these models were visualized, illustrating the true positive, true negative, false positive, and false negative predictions.

Conclusion:

In conclusion, this project successfully addressed the challenge of predicting bank churn through a multi-step process involving data analysis, preprocessing, dimensionality reduction, and machine learning. The F1 scores achieved demonstrate the models' ability to effectively identify potential churners. Future work may involve hyperparameter tuning and exploring ensemble methods to further enhance model performance.

Code Snippets

Data Transformation and SMOTE:

```
# Apply transformations to address skewness

transformed_data = apply_transformations(data)


# Implement SMOTE for class imbalance

from imblearn.over_sampling import SMOTE

smote = SMOTE(random_state=42)

X_resampled, y_resampled = smote.fit_resample(transformed_data, target)
```

One-Hot Encoding:

```
# Perform one-hot encoding on categorical features

encoded_data = pd.get_dummies(data, columns=categorical_features)
```

PCA and Visualization:

```
# Apply PCA for dimensionality reduction

from sklearn.decomposition import PCA

pca = PCA(n_components=6)

reduced_data = pca.fit_transform(encoded_data)


# Visualize reduced-dimensional data

plt.scatter(reduced_data[:, 0], reduced_data[:, 1], c=target, cmap='coolwarm')

plt.xlabel('Principal Component 1')

plt.ylabel('Principal Component 2')

plt.title('PCA Visualization of Churn Data')

plt.show()
```

Model Training and Evaluation:

```python
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.svm import SVC
from sklearn.metrics import f1_score, confusion_matrix

# Initialize and train models
rf_model = RandomForestClassifier()
adaboost_model = AdaBoostClassifier()
svm_model = SVC()

rf_model.fit(X_resampled, y_resampled)
adaboost_model.fit(X_resampled, y_resampled)
svm_model.fit(X_resampled, y_resampled)

# Make predictions
rf_predictions = rf_model.predict(X_test)
adaboost_predictions = adaboost_model.predict(X_test)
svm_predictions = svm_model.predict(X_test)

# Evaluate models
rf_f1 = f1_score(y_test, rf_predictions)
adaboost_f1 = f1_score(y_test, adaboost_predictions)
svm_f1 = f1_score(y_test, svm_predictions)

# Generate confusion matrices
conf_matrix_rf = confusion_matrix(y_test, rf_predictions)
conf_matrix_adaboost = confusion_matrix(y_test, adaboost_predictions)
conf_matrix_svm = confusion_matrix(y_test, svm_predictions)
```