AI-DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH REGISTRAR OF COMPANIES (RoC)

SUBMITTED BY
MANU JEEVA E
Dept. of Electronics and Communication Engineering
University College of Engineering Kanchipuram

Ensemble Methods:

- Ensemble methods involve combining the predictions of multiple machine learning models to improve overall performance.
- Techniques like Random Forests, Gradient Boosting, or stacking models can be employed to create a more accurate and robust prediction system.
- These methods are particularly useful when individual models may have their weaknesses, and combining them can compensate for those weaknesses

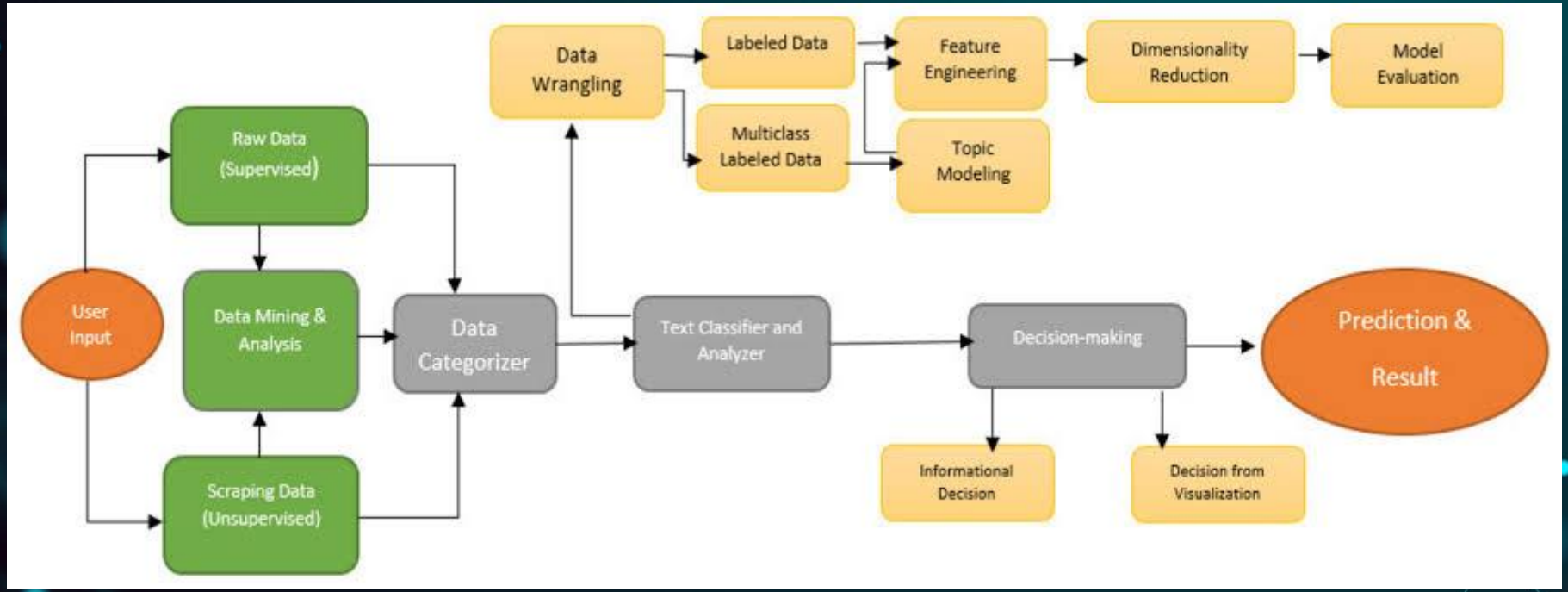
Deep Learning Architectures:

•Deep learning, a subset of machine learning, involves neural networks with many layers (deep neural networks).

• Deep learning architectures, such as convolutional neural networks (CNNs) for image data or recurrent neural networks (RNNs) for sequential data, can be explored to tackle complex prediction tasks.

•Deep learning has shown great promise in various fields, including natural language processing, computer vision, and speech recognition.

FLOW CHART


Data Wrangling
Labeled Data
Feature Engineering
Dimensionality Reduction
Model Evaluation
Multiclass Labeled Data
Topic Modeling
Raw Data (Supervised)
User Input
Data Mining & Analysis
Data Categorizer
Text Classifier and Analyzer
Decision-making
Prediction & Result
Informational Decision
Decision from Visualization
Scraping Data (Unsupervised)

Time Series Forecasting:

•If your prediction system deals with time-dependent data, time series forecasting methods can be incredibly valuable.

•Techniques like ARIMA (AutoRegressive Integrated Moving Average), LSTM (Long Short-Term Memory), or Prophet can help you make accurate predictions for future time points.

• Time series forecasting is commonly used in fields like finance, economics, and weather prediction

CODE

UCEK_IBM_AI.py - C:/Users/jjerl/Desktop/project/UCEK_IBM_AI.py (3.11.5)

File Edit Format Run Options Window Help

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import cross_val_score
from sklearn.metrics import roc_curve, roc_auc_score
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier
import matplotlib.pyplot as plt

# Step 1: Data Collection
data = pd.read_csv('registration_data.csv')

# Step 2: Feature Engineering
X = data.drop('Registration', axis=1)
y = data['Registration']

# Step 3: Model Building
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Step 4: Model Evaluation
model = RandomForestClassifier()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Step 5: Threshold Selection
y_prob = model.predict_proba(X_test)[:, 1]

# Step 6: Validation and Testing
cv_scores = cross_val_score(model, X_train, y_train, cv=5, scoring='roc_auc')
roc_auc = roc_auc_score(y_test, y_prob)

# Step 7: Interpretation and Action
feature_importance = model.feature_importances_
# Analyze feature importance and draw insights.

# Step 8: Continuous Monitoring
# Schedule periodic retraining of the model

# Optional: Plot the ROC curve
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
plt.plot(fpr, tpr, label='ROC Curve')
plt.xlabel('False Positive Rate')
plt.yl
```

Ensemble Methods for Improved Predictive Accuracy

- Reiterating the importance of ensemble methods, combining various models, even from different algorithm families, can lead to improved predictive accuracy.

- Combining decision trees with neural networks or linear models with support vector machines, for example, can result in a more robust and accurate system.

CONCLUSION

- In closing, our exploration of AI-driven analysis and prediction of company registration trends with the Registrar of Companies (RoC) has illuminated the immense potential that lies at the intersection of data, technology, and business strategy. Let's recap the key takeaways:

- Unlocking Insights: RoC data is a treasure trove of information that can provide valuable insights into the business landscape, economic trends, and industry growth.

- Data Quality Matters: The quality and integrity of data are paramount. Proper data collection, cleaning, and feature engineering are the foundation of meaningful analysis.

- AI's Predictive Power: Machine learning models, such as ARIMA and LSTM, enable us to forecast future registration trends with precision, helping stakeholders make informed decisions.

THANK YOU