Security Services Operations (SSO) Project Documentation

About SSO

SWIFT (Society for Worldwide Interbank Financial Telecommunications) is a global messaging network and financial communication platform used by banks and financial institutions worldwide. It is crucial for the majority of international money and securities transfers, providing a quick, accurate, and secure method for banks and financial institutions to send and receive information, such as money transfer instructions.

Problem Statement/Objective

The SWIFT network processes and routes messages to the recipient's bank (the beneficiary bank) using the SWIFT code mentioned in the message. These messages may also pass through intermediary banks, depending on the complexity of the payment route. Currently, several people are involved in processing, analyzing, and rerouting these messages. Our objective is to reduce the manual effort involved and automate this process using machine learning (ML), thereby providing a significant efficiency gain equivalent to 23 Full-Time Equivalents (FTEs).

Data Overview

- **Volume and Source:** 8,000 SWIFT messages received over 6 months from the Operations team.
- **Geographic Coverage:** Messages from MU, ZA, KE, AE, DC, and TH countries.
- **Message Types:** MT540 (Receive Free of Payment), MT541 (Receive against Payment), MT542 (Delivery against Payment), MT543 (Delivery vs Payment).
- **Entities to Extract:** 23 entities including Sender BIC, Trade Date, Settlement Date, Order Type, Function (Amend/Cancel), Client Reference, Safekeeping Account, Cash Account, Buy/Sell BIC, Buy/Sell Account, Deal Price, Currency, Security ID, Security No., Counterparty BIC 1, Counterparty BIC 2, CP Account, PSET Date, Quantity, Settlement Amount, Quantity Qualifier, Deal Price Currency, and IsAdjustment.
- **Tag 79:** A free-form message where users can write anything without following a particular standard.
- **Classification Categories:** 12 classifications including Settlement, Corporate Action, Billing, Trade Capture, Trade Capture Amend, Trade Capture Cancel, Account Management, MFA, AE (Control and Support), DC (Control and Support), FMO, Sanctions, Undefined.

Exploratory Data Analysis (EDA)

- **Imbalance:** There is a significant class imbalance among the 12 classification categories.
- **Dominant Categories:** The majority of messages fall under the Settlement or Corporate Action (CA) categories.

Preprocessing Steps

1. **Stop Words Removal:** Eliminate common stop words to reduce noise in the data.
2. **Pronoun Removal:** Exclude pronouns to focus on meaningful content.
3. **Country Name Removal:** Filter out country names to prevent geographic bias.
4. **Proper Noun Removal:** Remove proper nouns to enhance generalization.
5. **Lemmatization:** Normalize words to their base or root form using lemmatization.
6. **Vectorization:** Apply vectorization using SpaCy embeddings to convert text into numerical representations.

Classification

- **Algorithm Selection:** A custom model/algorithm selection framework was used to identify the **Random Forest Algorithm** as the optimal choice.
- **Performance Metrics:** The model's performance metrics, including accuracy, precision, recall, and F1-score, were thoroughly evaluated.

Entity Extraction

- **Trade Capture Class:** Custom Python scripts were developed to extract 23 entities specifically for the Trade Capture class.
- **Data Integration:** The extracted entities are realigned into a DataFrame and converted into JSON format for routing to the appropriate teams.

Sanity Testing

- **Cross-Country Validation:** The model is validated using random subsets of data from different countries (MU, ZA, TH) to ensure robustness.
- **Regression Analysis:** Any regression issues are identified and addressed in collaboration with the Operations team.

User Acceptance Testing (UAT)

- **Model Validation:** The Model Validation team tests the model with new or sequential SWIFT messages post-deployment.
- **Explainability:** Misclassifications are analyzed using LIME (Local Interpretable Model-Agnostic Explanations) to understand and interpret the model's predictions.
- **Retraining:** Feedback from LIME is used to retrain the model, ensuring continuous improvement and alignment with operational needs.

System Integration Testing (SIT)

- **Compliance Evaluation:** Comprehensive testing is conducted on the integrated system to ensure it meets specified requirements.

Model Deployment and Validation

- **Deployment:** Models are exposed as REST APIs using the Django framework, containerized with Docker, and deployed on OpenShift.
- **CI/CD Pipeline:** Azure DevOps is utilized for Continuous Integration/Continuous Deployment (CI/CD) and automated testing.
- **Ongoing Validation:** Continuous UAT ensures the model remains accurate and effective in production.

Benefit

- **Efficiency Gain:** The automation initiative has resulted in an efficiency gain equivalent to 23 Full-Time Equivalents (FTEs), significantly reducing manual efforts and enhancing operational productivity.