Loan Servicing Email Classification Pipeline

Introduction

Overview of the Loan Servicing Email Classification Project

Objective: Classify loan-related emails into request types, sub-request types, and extract key attributes

Problem Statement

Loan servicing centers receive numerous customer emails

Emails contain varied request types and key information

Manual classification is inefficient and error-prone

Goal: Automate classification and extraction using AI

Data Collection & Preparation

Emails stored in .eml format

JSON file mapping emails to request types, sub-request types, and key attributes

Attachments (PDFs) for additional data extraction

Data Preprocessing

Extracting text from .eml files using Pythons email library

Extracting text from PDFs using pdfplumber

Cleaning and tokenizing text using spaCy

Extracting key attributes (loan numbers, dates, amounts) using Named Entity Recognition (NER)

Model Selection

BERT (Bidirectional Encoder Representations from Transformers) chosen for classification

Fine-tuned for text classification with request type and sub-request type labels

Data Encoding & Tokenization

Label encoding for sub-request types using Scikit-learn

Tokenization of email text using BERT tokenizer

Loan Servicing Email Classification Pipeline

Dataset Splitting

Training and testing split (80%-20%)

Custom PyTorch Dataset class for handling tokenized inputs

Model Training

Using Hugging Faces Trainer API

Training arguments:

- Learning rate: 2e-5
- Batch size: 8
- Number of epochs: 3
- Weight decay: 0.01

Evaluation strategy set to epoch-wise validation

Model Evaluation

Compute accuracy and classification report

Evaluating models performance on test dataset

Model Deployment

Saving trained model and tokenizer

Future integration with APIs or backend systems

Challenges & Improvements

Handling multiple requests in a single email

Improving accuracy with more data and hyperparameter tuning

Better handling of attachments for key information extraction

Conclusion

Successful automation of loan email classification

Reduces manual effort and increases efficiency

Loan Servicing Email Classification Pipeline

Future scope: Deploying model into production and enhancing feature extraction

Thank You!