

BSA Data Mining

Building a Knowledge Base for an Artificial Intelligence Chatbot

Joseph Bryant, Tyler Heald, Katherine Merola, An Nguyen, Jonathan White
jmb160330@utdallas.edu; Tyler.Heald@utdallas.edu; Katherine.Merola@utdallas.edu; An.Nguyen10@utdallas.edu;
Jonathan.White1@utdallas.edu;



CS 4485 / Spring 2020
Department of Computer Science
Erik Jonsson School of Engineering & Computer Science
The University of Texas at Dallas
Richardson, TX 75080, USA



Prepared. For Life.®

Abstract

Create a knowledge base of information to aid in identifying user issues in order to target the correct resolution to the problem. Past support tickets will provide labeled training data of user issues. This data will have to be cleaned and parsed before a classifier model is trained on the data. This classifier will be able to evaluate new user issue descriptions and predict the category that the issue belongs to. This will enable information from past support tickets to be used in directing new user issues towards a resolution. This classifier and the knowledge base it exposes will be able to interface with a web-based support chatbot. This will create an intelligent system where users can seek solutions to common issues.

Keywords: Data Mining, Artificial Intelligence, Machine Learning

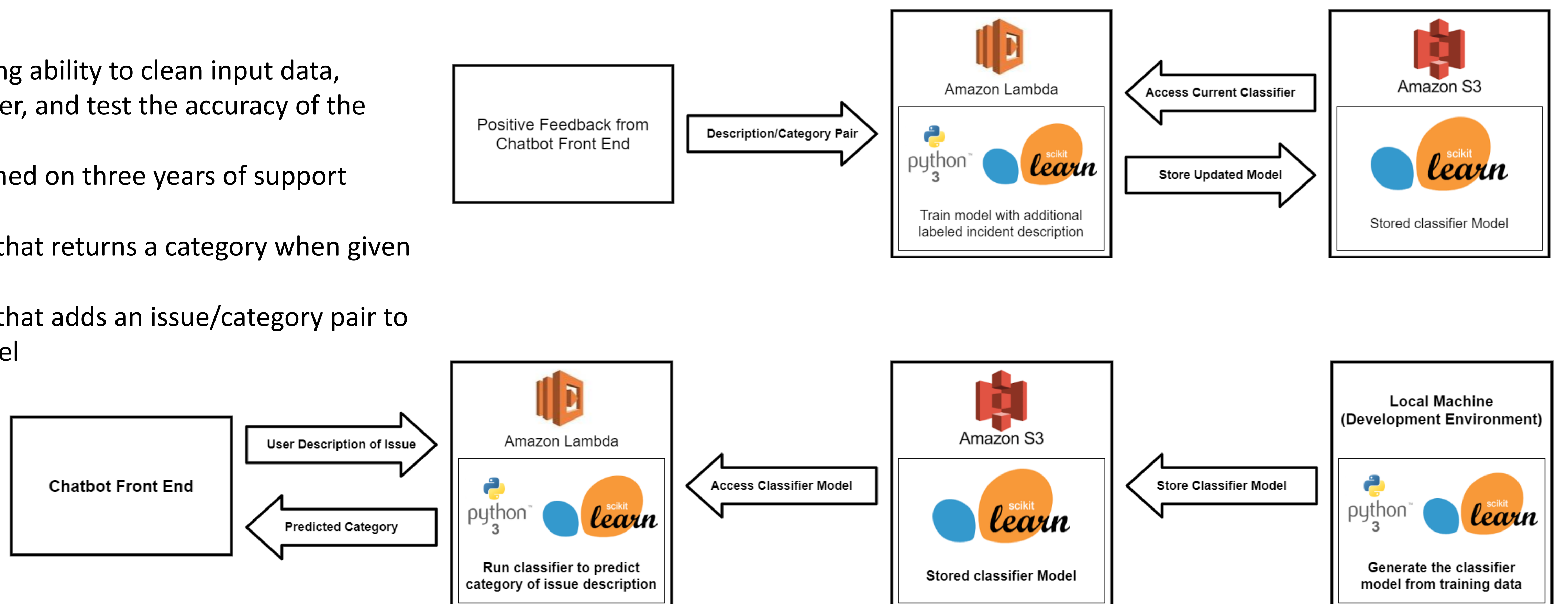
Architecture

- **Python3 Language and Libraries**
 - Scikit-learn – Machine Learning, classifier training
 - Natural Language Toolkit (nltk) – Text parsing
 - Pandas – Data analysis and manipulation
- **Amazon Lambda**
 - Execute code on demand
 - Integrates with other AWS services
 - Provides an API to the knowledge base
- **Amazon S3**
 - Secure file storage
 - Provides file access to other AWS services
 - Enables storage of trained classifier model



Results

- Codebase providing ability to clean input data, generate a classifier, and test the accuracy of the classifier.
- Text classifier trained on three years of support tickets
- Lambda function that returns a category when given a user issue
- Lambda function that adds an issue/category pair to the classifier model



Impact

The work on this project has several important effects:

- A classifier can be used to predict the category of a support issue described by a user.
- The classifier can be further trained and refined with additional user data, enabling a continually learning model.
- Ability to invoke these functionalities through AWS Lambda, allowing seamless integration with other AWS services.

The functionality of this project exposes the information encoded within three years of BSA support tickets, in order to build a condensed knowledge base, rich with valuable information. This knowledge base provides quick and easy access to this vault of information, in order to support the mission of an AI support chatbot to provide solutions to the user.

Performance

- On its own, the performance of this project can be measured by the accuracy of the trained model in predicting categories successfully.
- The most important performance of the project was that it could interface with and support the chatbot to provide intelligent identification of user issues. This performance was achieved with the development of an API in AWS Lambda.

Summary

- Built a text classifier model with supervised machine learning from years of support tickets
- Created an API with Amazon Lambda to interface with the classifier
- Provides a knowledge base to a support chatbot, creating an intelligent system