Accuracy Development Plan to Reduce False Positives in Reputation Checking

Prashant Moré, MS Finance 2020-21

Public Summary Report

Contents

1   Executive Summary

This document describes proposed developments to enhance the efficiency of the bad news engine. The company's existing code was giving false positives when reputation of an individual is checked using bad news search engine. In this capstone project, it was proposed, implemented and demonstrated that using additional keywords with Google search operators reduces the number of False Positives.

2   Client Background

The company has developed an Artificial Intelligence (AI) enabled bad news engine for small and medium businesses looking to mitigate their risks. The mission is to help small teams make informed business decisions about their customers without spending a fortune on due diligence. It will be achieved by capitalizing on machine power to increase the speed and accuracy of information processing. (Source: the company website)

3 Issues Addressed

This document addresses the problem the company is currently facing, that is too many False Positives (FP). FPs are bad as they increase the runtime, decrease the efficiency of the engine, and possibly may lose customers as they find the results inappropriate. There is also the possibility of client taking an adverse decision, resulting in a possibility of litigation.

4 Project Timeline

The project was started in the last week of March 2021. The kick-off documents were submitted to CEU on 7th April 2021. The interim report was due on 10th May 2021. The final deliverables were submitted to the company on 4th June 2021, due on 8th June 2021 to CEU.

5 Deliverables

- Website for the company developed using Bootstrap template.

- Custom modifications done as applicable to the company.

- MVP developed which mimics bad news engine functionality.

- Demonstration of reduction in False Positives.

- Admin area developed and superusers added to test generated code in Django framework.

6 Conclusions and Lessons Learnt

1. Documentation of work done for any purpose in a company is very important as it allows others to carry foreword the work completed. Also, it helps understand the logic of the previous person's work.

2. Django is a framework of Python which enhances its usability by allowing developers in creating user interaction elements. Python itself is a maturing programming language with ever growing library designed for various purposes such as data visualization (Matplotlib), natural language processing (NLTK and Spacy), google search (Google-search), customizing query and search elements (Regular Expressions) and data processing (Panda).

3. The company's website was also developed in Django using bootstrap 4 template. In Django, the static files are extremely important for correct rendering of the website. Also, at the time of deployment, the debug variables have to be changed and allowed hosts have to be modified.

4. Localhost serves extremely efficient way of testing websites in a local environment. On a Mac, terminal allows execution of python commands in a shell environment. This is useful because databases can be queried without building user interfaces.

5. Google search operators play an important role in bad new search engine applications. Search engine query can be enhanced, modified by applying operators either with logical keywords or special symbols.

6. Github plays very important role, if you are submitting the code to someone for review asking his opinion. Github nicely integrates with Visual Studio Code for one-click importing of a repository. It also allows to commit changes to the repository.

7. Proper indentation in coding is an absolute must. In wsgi files, whitespaces are crucial and can break the code. Comments at appropriate location are required, for the developer him/herself, as it is easy to forget the logic later on.

8. For simple applications, Django is an overkill. Flask can be substituted. However, in current project, existing work was done in Django, hence Flask implementation was not possible.

9. Amazon Web Services (AWS) provides one stop cost-effective services for most of the web development needs. EC2, Lightsail, ElasticBeanstalk and S3 services were explored and used for this project. Alternative Services such as Heroku and Digital Ocean can be also explored in future.

10. Final report writing should be concurrent with code development. Certain libraries made specifically for Latex typesetting allows easy code formatting. Flowcharts are rendered effectively in Latex.

11. Python can automate a lot of corporate procedures used in Finance effectively. Since it has libraries related to Statistics, visualization, data analysis and other scientific computing, there is a huge untapped market for corporate level consulting. It is possible that in most companies, senior level management is not yet potentially aware about possibility of implementing automatic state-of-the-art technology especially in financial reporting, accessing and processing stock prices, algorithm trading, risk management procedures, etc.