

Axis Bank Innovation Hack

Software Design Document

College of Engineering Pune

Team members:

Ashish Kulkarni: kulkarniashish88@gmail.com

Saket Komawar: saketkomawar95@gmail.com

Shubham Shah: shubhamshah970@gmail.com

TABLE OF CONTENTS

Sr. no	Title	Page no.
1	INTRODUCTION	iii
1.1	Scope	iii
2	SYSTEM OVERVIEW	iv
2.1	System Characteristics	iv
2.2	System Architecture	v
3	COMPONENT DESCRIPTION	vi
4	SOFTWARE DEVELOPMENT TOOLS	vii

1 INTRODUCTION

This document describes a system which helps predict whether particular business units are in need of foreign exchange services. This system can be of great significance to units like banks which provide foreign exchange services in helping them (banks) predict their customers. The system leverages machine learning, natural language processing, search engine APIs and web scrapping tools to achieve the end goal of predicting.

The system is a web application hosted on a cloud server. All authorized users across various locations (e.g. employees of banks working in various branches across the globe) can access it. The system takes as input the name of company or business unit and in return generates the propensity score (probability) of whether that company is in need of foreign exchange services. The system using Bing search engine to gather information about that company, scraps the information from the internet, applies natural language processing techniques on it to extract features and finally feeds the information to a machine learning model which calculates the propensity score.

1.1 SCOPE

Scope of the system is as follows:

- a. Predict whether a company will require foreign exchange services
- b. User of the system need not install any software to be able to use it as it is a web application and can be accessed using any browser
- c. If newer features come up in future for the prediction task, the system can easily accommodate the new features with minimal changes
- d. System can handle high work loads owing to the scalability features of the cloud provider

2 SYSTEM OVERVIEW

The code base of the system is mostly written in Python. The system is basically a web application developed using Python's Django web framework. System uses Bing search engine tools to get information about a particular company. Microsoft Cognitive Services API are used to access the Bing search engine. BeautifulSoup web scrapper is used to scrap text information from the web pages returned by Bing. Natural Language Toolkit (NLTK) is used for natural language processing of the text returned by web scrapper. Xgboost which implements gradient descent and tree boosting machine learning techniques is used to implement the machine learning model.

2.1 SYSTEM CHARACTERISTICS

Following are the features extracted from text information which are used to build the machine learning model and to do predictions:

1. Importer Exporter Code (IEC) of the company

The system has a database of all IEC holder companies. It checks if the company under consideration has its IEC. If it has its IEC it is highly probable that the company may need foreign exchange service

2. Number of country names found in the text information of company

More the number of country names found in webpages like company's website, it signifies that either company has offices overseas or has customer overseas, in both cases, it is probable that company would need foreign exchange services.

3. Number of international city names found in the text information of company

The motive behind using this feature is similar to that of number of country names

4. Count of occurrence of keyword 'import' in the text information

If keywords like 'import' are mentioned in the company information data, it is highly likely that the company imports some sort of goods or services. This increases the chances of company needing forex services.

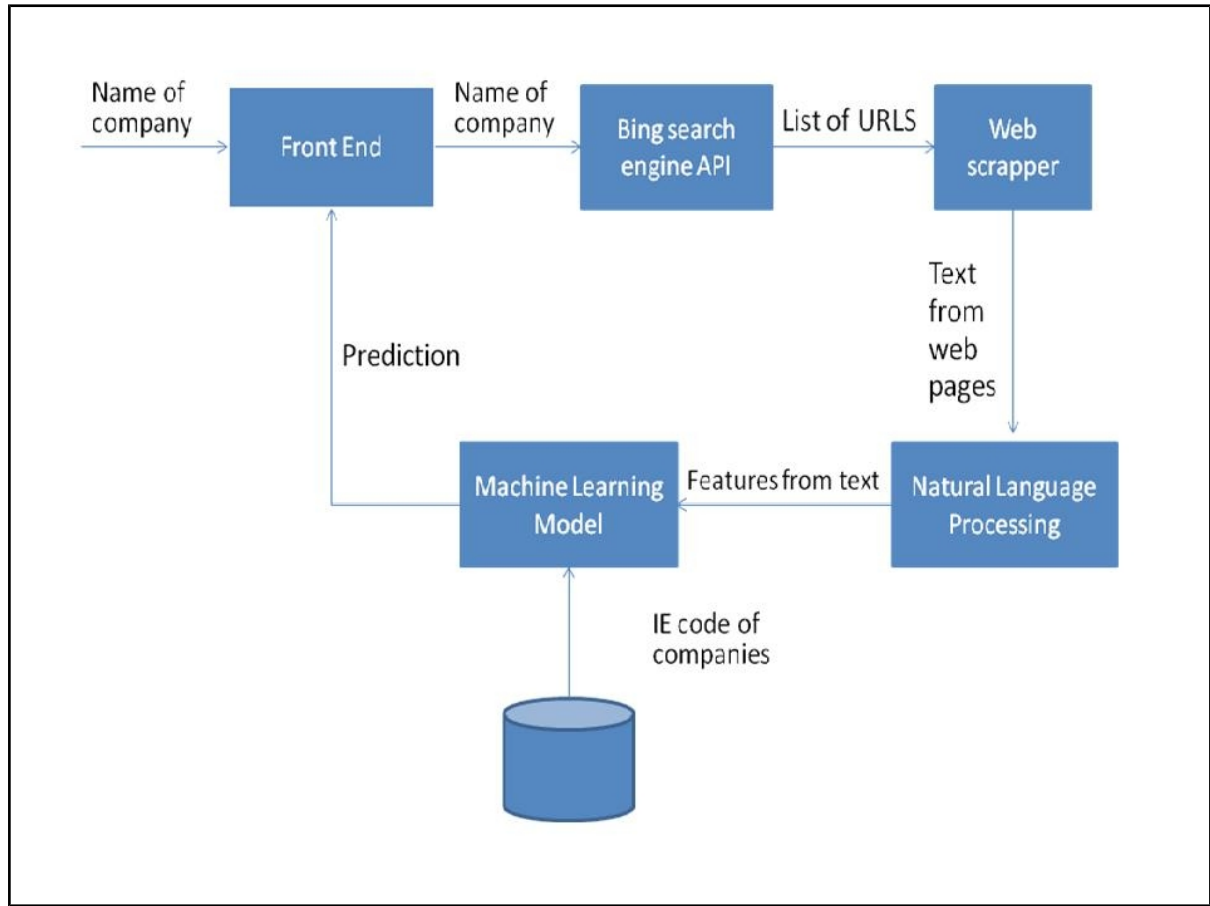
5. Count of occurrence of keyword 'export' in the text information

Mention of the keyword 'export' signifies that the company is involved in some sort of export, thereby raising the chances of needing forex services.

6. Count of occurrence of keyword 'foreign exchange' and 'currency exchange' in the text information

If the company information has mentions of keywords like 'foreign exchange' or 'currency exchange' it is highly likely that they need/use one.

2.2 SYSTEM ARCHITECTURE



3 COMPONENT DESCRIPTION

Following is the list of major blocks of the system and their brief explanation.

Front-end: This consists of the web application user interface. User can either enter name of a single company or upload a file containing list of company names whose need for forex services is to be predicted.

Bing search engine API: This block takes as input the name of company and returns the list of urls obtained after doing a bing search for the company name

Web scrapper: This block takes as input the list of urls returned by Bing and in return scraps (downloads) all the text information on these webpages.

Natural Language Processing: The input to this block is the text gathered by the web scrapper and output is the list of features extracted from the text. It used named entity recognition technique of natural language processing.

Machine learning model: This block takes as input the list of features, applies the machine learning algorithms on these features and gives as ouput the probability of the particular company needing forex services.

4 SOFTWARE DEVELOPMENT TOOLS

The following is a list of various software development tools and their use in the system.

Python: Major programming language for the system

HTML: Used for the designing the web pages

CSS: Used for styling the web pages

Django Web Framework: Used to create the web application

Microsoft Cognitive Services: APIs to interact with Bing search engine

Natural Language Toolkit (NLTK): For natural language processing, particularly entity recognition

Xgboost: Machine learning library

Beautiful Soup: Web scrapper to get information from web pages

Postgres SQL: Database for the system