Risk Identification and Assessment Document

Project: Sentimento

"Social media assisting platform with sentiment analysis"

Table of Contents

Introduction

This documentation covers the risk/sensitive information involved areas of the 'Sentimento' application. This document is the process of determining all the possibility or likelihood of any unwanted scenarios that might occur in the future if bad actors are involved in the platform.

Apart from user data misuses, this document also includes the challenges to built a functioning sentiment analysis platform.

Purpose

Primary objective of this document is to provide substantial information about challenges involved in building this application. This document clearly highlights all those scenarios. The other tool and technologies requirements for the development of the project is also defined in this document.

The core purpose of this mobile application is to provide a user-friendly platform where non-technical and technical (with understanding of Data Analytics) users can perform sentiment analysis within same interface and connect job seekers with open vacancies.

Intended Audiences of this document

This document is primarily intended for the development team of the Sentimento project. Developers from algorithm team, backend and frontend including other team members engaged in technical and non-technical areas are the key readers/viewers of this document. The client of the project or potential users can also read this document to verify their demands are included in the document.

Why this document for the developer team?

Viewing all the features to include in the platform, developers can identify and plan their moves accordingly. The vocal hearing from the client is not consistent and official but this document with client's approval provides clear vision on what to do and how to implement those essential features in the platform. This document can also be a proof of record of tasks to carry out for both the parties (client and development team).

Professional Issues

Here's how the platform is built to prevent any professional issues from happening:

Creating a product with the potential of numerous users to be involved adds a sensitive responsibility to protect the rights and information of consumers of the product. Considering this part of the professional ethics and computer law: certain ethics, moral and legal duties are strictly followed in the development of this project which are as follows:

i.   Users are trusting with their personal information while using this app, understanding this, passwords are hashed before storing in database. The email that user provides while registering in the platform is not visible to third person/parties.

ii.   The third party's (YouTube and Twitter) API key user has to enter to perform sentiment analysis are user's sensitive information. Keeping this in mind, those API keys are not stored in system's database. Instead, the keys are encrypted and locally saved in the user's device.

iii.   To prevent bad actors from posting spam vacancies, the system throttles users to post only five open vacancies at a time.

iv.   There might be any incorrect information while posting vacancy by user. In such case user wants to remove such misinformation. Considering this, there is a feature to delete previous posts.

Limitations

Challenges that can arise while performing sentiment analysis are listed below:

Sentiment Analysis is a complex and arduous task in natural language processing. The challenges of machine learning based sentiment analysis are as follows:

i.   Irony and Sarcasm: People uses sarcasm and ironic sentences to express the meaning. In such cases, the real point they want to make can be one thing but the straight sentence can mean something different. Even human can find it difficult to understand sometimes which makes more difficult for trained models to predict the actual meaning.

ii.   Context and Polarity: As trained model works within the provided boundaries, it's hard to understand the actual sentiment according to the context. Example: the opinion can be regarding something but the context can be different. Here the calculated sentiment polarity does to belong to real context resulting less trustworthy analysis.

iii.   Stop words: The stop words are better to remove for reducing the numerical vectorized array of data. This can save memory and better performance but removal of stop words can alter the meaning of a sentence. Example: In the sentence 'I do not like you'. When 'not' stop word is removed while pre-processing, the meaning changes. So, it is challenging to deal with stop words.

Limitations of this built application:
The designed algorithm is not well optimized to perform analysis quickly. There is still adequate place to improve time and space complexity of the algorithm. So, the platform is throttled to only take maximum of one thousand data (comment/tweet) at a time.