**FINE MINE**

**RESUME MINING**

**INTRODUCTION**

Majority of world’s unstructured data is in the textual form. To make sense of it, one must, either go through it painstakingly or employ certain automated techniques to extract relevant information. Looking at the volume, variety and velocity of such textual data, it is imperative to employ Text Mining techniques to extract the relevant information, transforming unstructured data into structured form, so that further insights, processing, analysis, visualizations are possible.

This document deals with a applicant profiles or resumes. They, as we know, come not only in different file formats (txt, doc, pdf, etc.) but also with different contents and layouts. Such heterogeneity makes extraction of relevant information, a challenging task. Even though it may not be possible to fully extract all the relevant information from all the types of formats, one can get started with simple steps and at least extract whatever is possible from some of the known formats.

Broadly there are two approaches: linguistics based and Machine Learning based. In “linguistic” based approaches pattern searches are made to find key information, whereas in “machine learning” approaches supervised-unsupervised methods are used to extract the information. “Regular expression” (RegEx), used here, is one of the “linguistic” based pattern-matching method.

A primitive way of implementing entity extraction in a resume could be to write the pattern-matching logic for each entity, in a code-program, monolithically. In case of any change in the patterns, or if there is an introduction of new entities/patterns, one needs to change the code-program. This makes maintenance cumbersome as the complexity increases. To alleviate this problem, separation of parsing-logic and specification of entities is proposed in a framework, which is demonstrated below. Entities and their RegEx patterns are taken care.

**APPLICATIONS OF TEXT MINING**

Text mining techniques and text mining tools are rapidly penetrating the industry, right from academia and healthcare to businesses and social media platforms. This is giving rise to a number of text mining applications. Here are a few text mining applications used across the globe today:

1. **Risk Management**

One of the primary causes of failure in the business sector is the lack of proper or insufficient risk analysis. Adopting and integrating risk management software powered by text mining technologies such as SAS Text Miner can help businesses to stay updated with all the current trends in the business market and boost their abilities to mitigate potential risks. Since text mining tools and technologies can gather relevant information from across thousands of text data sources and create links between the extracted insights, it allows companies to access the right information at the right moment, thereby enhancing the entire risk management process.

1. **Customer Care Service**

Text mining techniques, particularly NLP, are finding increasing importance in the field of customer care. Companies are investing in text analytics software to enhance their overall customer experience by accessing the textual data from varied sources such as surveys, customer feedback, and customer calls, etc. Text analysis aims to reduce the response time of the company and help address the grievances of the customers speedily and efficiently.

1. **Fraud Detection**

Text analytics backed by text mining techniques provides a tremendous opportunity for domains that gather most data in the text format. Insurance and finance companies are harnessing this opportunity. By combining the outcomes of text analyses with relevant structured data these companies are now able to process claims swiftly as well as to detect and prevent frauds.

1. **Business Intelligence**

Organizations and business firms have started to leverage text mining techniques as part of their business intelligence. Apart from providing profound insights into customer behaviour and trends, text mining techniques also help companies to analyse the strengths and weaknesses of their rivals, thus, giving them a competitive advantage in the market. Text mining tools such as Cogito Intelligence Platform and IBM text analytics provide insights on the performance of marketing strategies, latest customer, and market trends, and so on.

1. **Social Media Analysis**

There are many text mining tools designed exclusively for analyzing the performance of social media platforms. These help to track and interpret the texts generated online from the news, blogs, emails, etc. Furthermore, text mining tools can efficiently analyze the number of posts, likes, and followers of your brand on social media, thereby allowing you to understand the reaction of people who are interacting with your brand and online content. The analysis will enable you to understand ‘what’s hot and what’s not’ for your target audience.

**IMPORTANCE OF TEXT MINING IN DATA MINING**

Data and information have grown at an amazing rate due to the quick increase of computerized or digital information. Text databases, which include enormous collections of documents from diverse sources, are where a significant amount of the information that is now available is kept.

Due to the enormous amount of information available in electronic form, text databases are expanding quickly. Over 80% of the knowledge available today is unstructured or somewhat loosely arranged. The growing volume of text data makes outdated information retrieval methods ineffective. As a result, text mining is now a crucial and widely used component of data mining. In practical application domains, identifying appropriate patterns and analysing the text document from the enormous volume of data is a significant challenge.

**Industries that use Text Mining efficiently –**

1. **Financial Services**

The financial services industry is incredibly intricate. It involves a significant quantity of communication, paperwork, risk assessment, and compliance. Financial services companies use text analytics to examine client comments, assess claims, consider consumer interactions, and pinpoint compliance issues. Staff members may quickly and easily search internal legal papers for terms related to money or fraud using a text analytics system built on NLP. When compared to complete it manually, this can save a significant amount of time.

1. **Healthcare and Pharma**

Specialists in medical affairs assist in the transition of pharmaceutical goods from R&D to commercialization. Text mining is being used by experts in medical affairs to automatically interpret each of these and report changes. Depending on what these alterations indicate for the medication they are creating, the specialists can then adjust their direction. Instead of relying on human labor, text mining can track these changes more accurately and extensively while taking up less time.

1. **Retail**

The consumer is always correct in the retail industry. With the surge in online sales during the pandemic, e-Commerce sellers must make sure that the consumer experience is as favourable as possible. Even more so than at physical establishments that people visit, a bad experience makes a client unwilling to return. Text mining is being used by many e-tailers to collect, organize, and analyse consumer input that identifies places of friction while using an e-commerce website or interacting with customer care.

**APPLICATIONS OF RESUME MINING**

* Risk Management
* Knowledge Management
* Cybercrime Prevention
* Enhanced Customer Service
* Streamlined Claims Investigation
* Contextual Advertising
* Business Intelligence
* Content Enrichment

**SOFTWARE REQUIREMENTS**

**VS CODE**

We use Visual Studio Code for running all our code. Where we can run our frontend and backend code easily. Giving connection between frontend and backend will be convenient using VS code.

**Mongo DB**

As we are extracting dynamic data from the resume, we are using MongoDB (NoSQL) which will help us to store our data, retrieve it very easily.

**Front End**

Here we are using Html, CSS for frontend. In html pages we also use Jinga Tags that will help us to use python code inside the Html pages. And we are using CSS for a better UI for the user.

**Back End**

For backend we used python as coding language and we use flask frame work to create our application. We used PyMongo for database connection with MongoDB.

**Database**

MongoDB is a NoSQL database which will store the data in unstructured format.

We are extracting data from the resume and saving it into the database for the further reference.

**WORKFLOW**

When we send or upload a resume, we will convert the resume into binary format and extract all the skills from the resume. Features are:

* Skill extraction
* Extracting programming language
* Web Development, Database Query Language
* Tools
* Scripting Language
* Certification
* Project

As we started building resume mining application for the IT recruitment, we have mentioned those categories for getting the data. But based on the need we can able to change the categories. That will be help full for developing the application into more futuristic one.

We store every data we extract from the resumes that will help us prevent spam on the application that save more amount of time. This will also help us to differentiate the candidate and able to make future move based on their resume itself. This will be saving a major amount of time for us and this will show us only the required candidates.

By using this application, the major goal will be to segregate candidates based upon company’s requirement and the sets skills by the candidate whichever criteria matches only those resumes will be highlighted.

**SCREENSHOTS**



Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

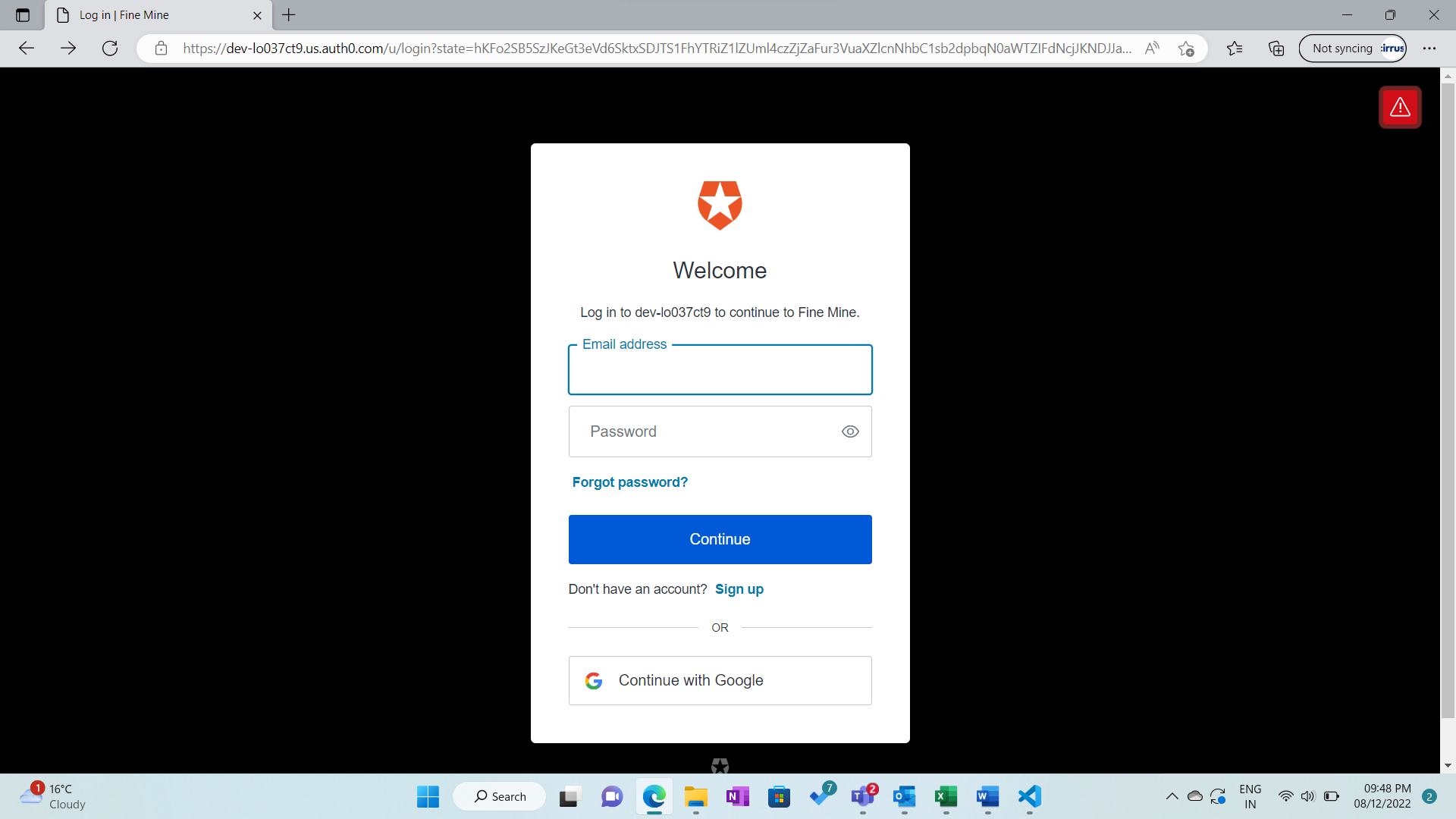
Description automatically generated

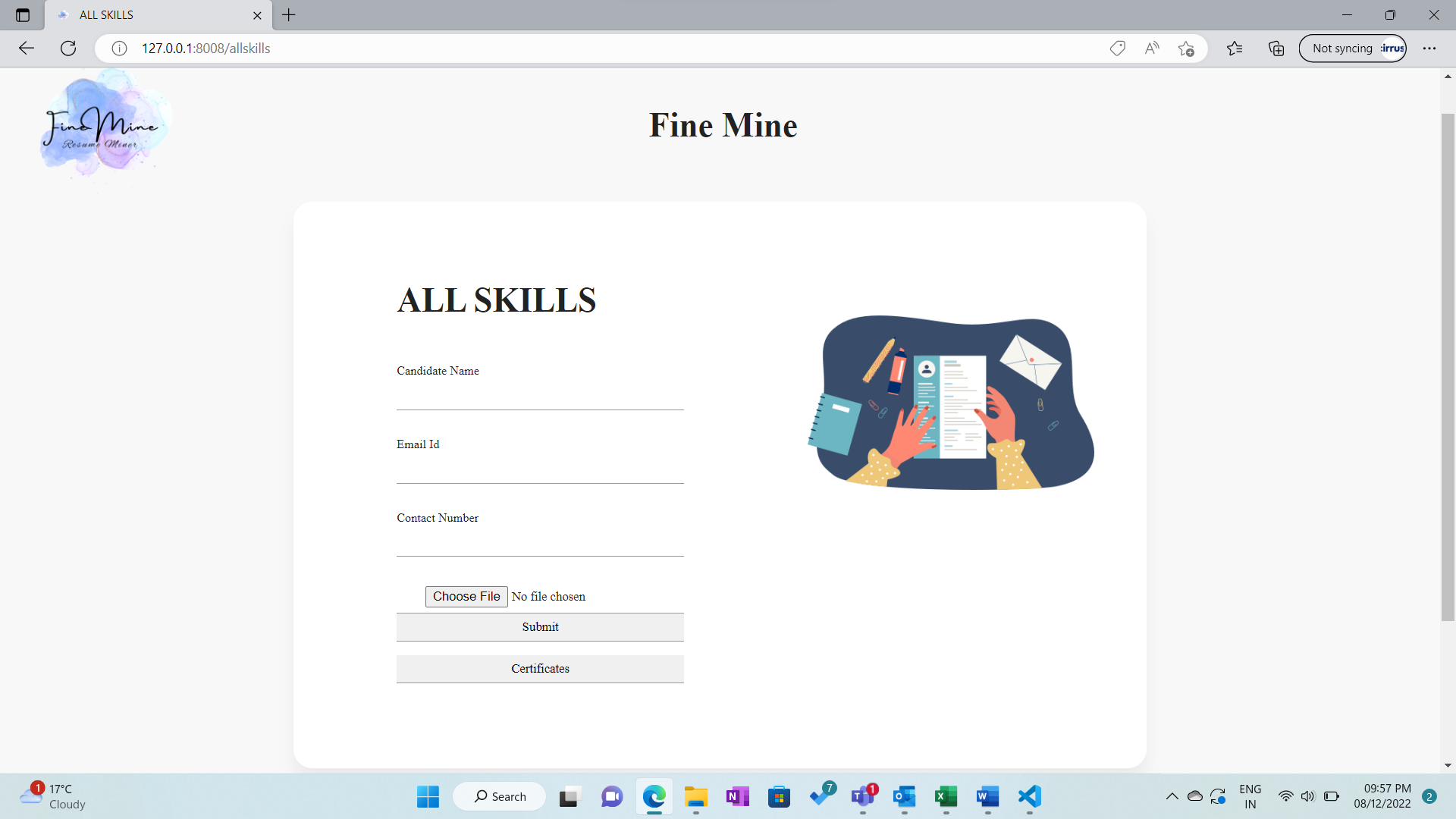
Graphical user interface, application

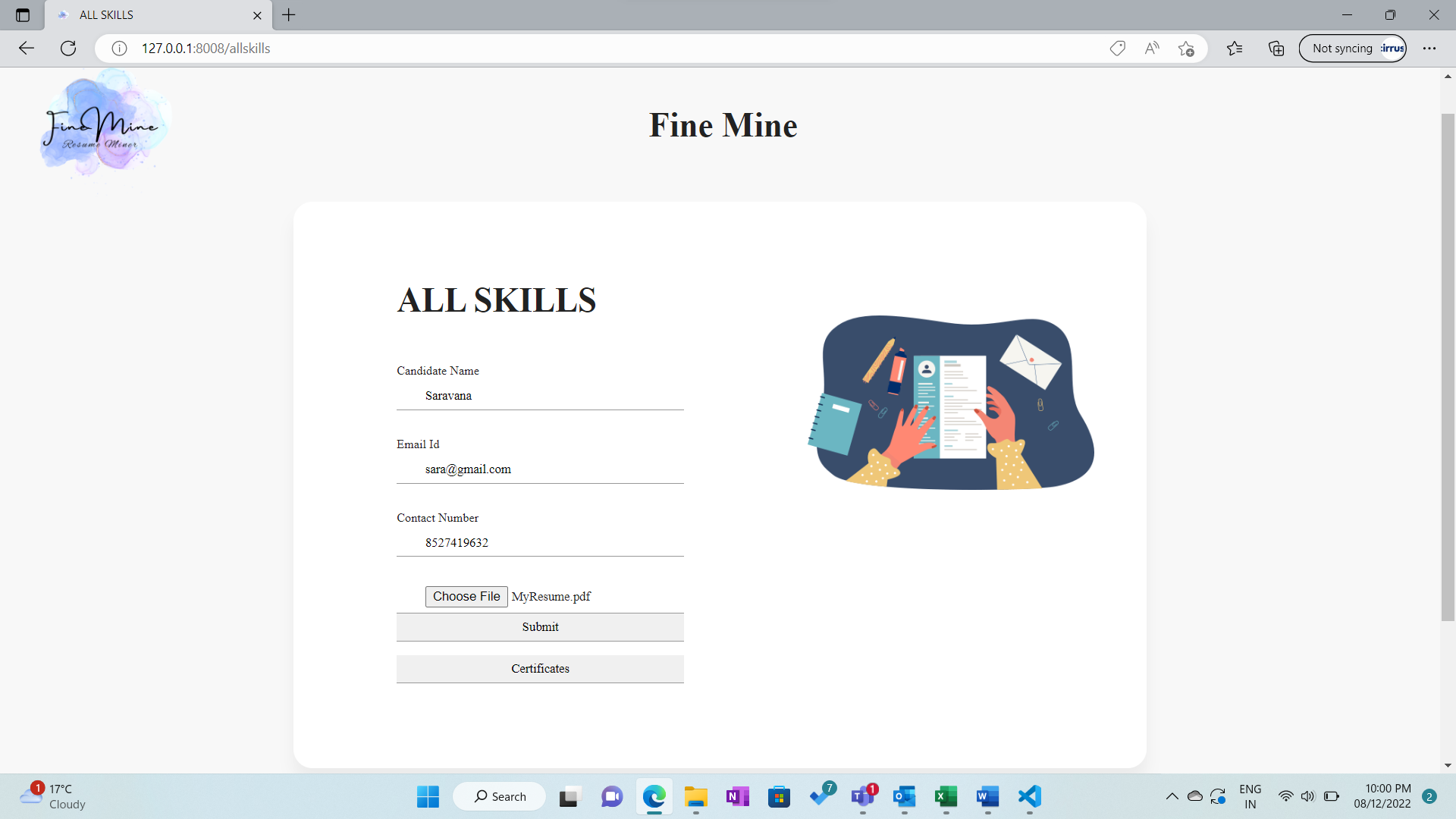
Description automatically generated

Graphical user interface, application, Word

Description automatically generated





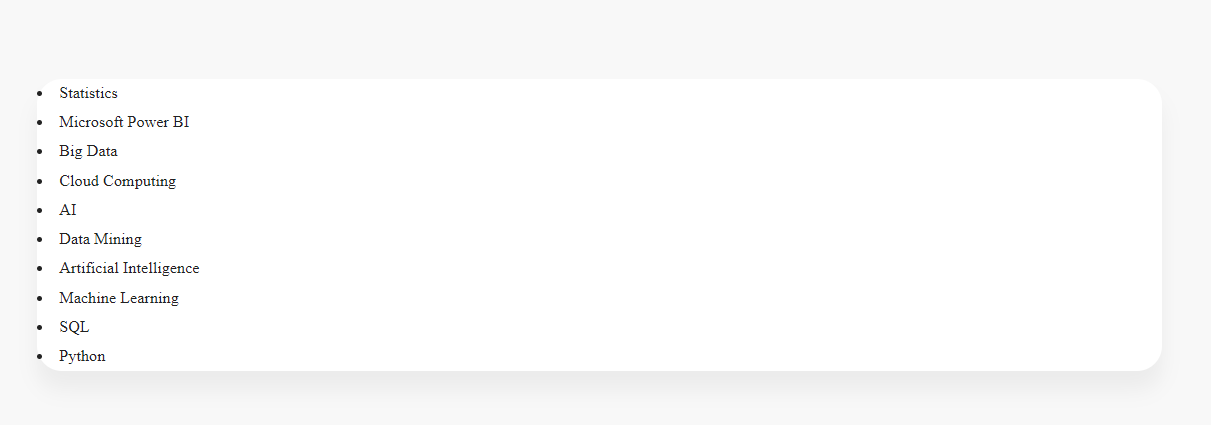


Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generated



Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**FUTURE ENHANCEMENT**

As of now we can able to select the candidates based on their skills and will filter the resume and give us needed resume. But it will not show that how much the candidate if preferable for the particular role. So, we were going to include this function by evaluating the certification, based on the project they have involved, based on the experience and their college we can say how much they are suitable for the particular job. Those are the things we were going to work on future.