माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA
Deemed to be University
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

A Skill Based Mini Project Report
on

"Question & Answer Based System Tool"

Submitted by

Pratham Bajpai (0901EO211043)

Submitted to

Dr. Murli Manohar
Assistant Professor

Centre for Internet of Things
Madhav Institute of Technology & Science, Gwalior
Gole ka Mandir, Gwalior - 474005, M.P., India

Jan – Jun 2024

Madhav Institute of Technology & Science, Gwalior

(Deemed to be University)

NAAC Accredited with A++ Grade

(Declared under Distinct Category by Ministry of Education, Government of India)

Centre for Internet of Things

DECLARATION

I/We hereby declare that the work being presented in this skill based mini project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Internet of Things at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr. Murli Manohar**, Assistant Professor, Centre for Internet of Things.

I/We declare that I/We have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Pratham Bajpai

(0901EO211043)

Centre for Internet of Things

Madhav Institute of Technology & Science, Gwalior

(Deemed to be University)
NAAC Accredited with A++ Grade
(Declared under Distinct Category by Ministry of Education, Government of India)

Centre for Internet of Things

CERTIFICATE

This is certified that **Pratham Bajpai (0901EO211043) and Khushi Kumar Rathore (0901EO211032)** has submitted the skill based mini project report titled **"Question & Answer Based System Tool"** under the mentorship of **Dr. Murli Manohar**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Internet of Things from Madhav Institute of Technology and Science, Gwalior.

23/05/2024.

Dr. Murli Manohar
Assistant Professor
Centre for Internet of Things

Madhav Institute of Technology & Science, Gwalior
(Deemed to be University)
NAAC Accredited with A++ Grade
(Declared under Distinct Category by Ministry of Education, Government of India)
Centre for Internet of Things

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology & Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Internet of Things,** for allowing me to explore this project. I humbly thank **Dr. Praveen Bansal**, Assistant Professor and Coordinator, Centre for Internet of Things, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Murli Manohar**, Assistant Professor, and Centre for Internet of Things, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Pratham Bajpai
(0901EO211043)
Centre for Internet of Things

TABLE OF CONTENTS

ABSTRACT

The Q&A System Based Tool developed for Codebasics represents a significant advancement in e-learning support services, leveraging advanced language models and AI technologies to provide students with quick and accurate responses to their queries. Through the integration of Google Palm, Langchain, and Streamlit, the system offers an intuitive user interface and seamless interaction, revolutionizing the way student questions are addressed within the platform. Key results demonstrate high efficiency, reduced workload for human staff, improved student satisfaction, and scalability for future growth. This abstract encapsulates the system's success in enhancing the learning experience for Codebasics students and sets the stage for further advancements in e-learning support services.

In the dynamic landscape of equity research and financial analysis, timely access to relevant information is paramount for making informed investment decisions. The News Research Tool represents a significant advancement in this regard, offering analysts a comprehensive solution for extracting insights from news articles in the stock market and financial domain. By harnessing the power of natural language processing (NLP) and artificial intelligence (AI), the tool automates the process of data collection, analysis, and retrieval, enabling analysts to focus their time and expertise on value-added tasks.

LIST OF FIGURES

Chapter 1:
Introduction

1.1 About Question & Answer System Based Tool

The Question & Answer (Q&A) System Based Tool is an innovative solution developed for Codebasics, an e-learning company specializing in data-related courses and bootcamps. With thousands of learners seeking assistance through Discord servers or email, Codebasics faces the challenge of efficiently addressing student queries and providing timely support. This tool aims to streamline the process by automating responses to common questions using advanced language models and AI technologies.

Key features of the (Q&A) System Based Tool include the ability to load article URLs or upload text files containing URLs, processing article content through LangChain's Selenium URL Loader, constructing embedding vectors using GoogleAI embeddings, and leveraging FAISS for efficient information retrieval. Additionally, users can interact with Gemini to ask questions and receive answers based on the analyzed news articles.

1.2 Parameters Considered For for Q&A Tools

In designing the Q&A system, several parameters were carefully considered to ensure its effectiveness and usability:

Accuracy: The system must provide accurate responses to student queries, drawing from a comprehensive knowledge base of FAQs.

Speed: Timely responses are crucial to student satisfaction. The system should deliver answers within seconds, enhancing the learning experience.

Scalability: As Codebasics serves a large number of learners, the system should be scalable to accommodate increasing volumes of queries without sacrificing performance.

User Interface: The system's interface should be intuitive and user-friendly, allowing students to ask questions easily and navigate the platform seamlessly.By considering these parameters, the News Research Tool aims to provide a comprehensive solution for equity analysts seeking to conduct research in the financial domain.

1.3 Objective of Project

The primary objective of the Q&A System Based Tool is to enhance the learning experience for Codebasics students by providing quick and accurate responses to their queries. By leveraging language models such as Google Palm and Langchain, the system aims to reduce the workload of human staff while ensuring that students receive the support they need in a timely manner. The project seeks to achieve the following goals:

1. Automation
2. Accuracy
3. Efficiency
4. Usability

Fig 1.1 Calling GoogleAPI

Chapter 2:
Literature Survey

In the realm of question and answer systems for e-learning platforms, various studies and publications have explored the application of natural language processing (NLP) and artificial intelligence (AI) techniques to enhance the learning experience. This literature survey delves into key research findings and methodologies relevant to the development of the Q&A System Based Tool for Codebasics.

1. **NLP in Education:**

   NLP techniques have been widely applied in educational settings to facilitate language learning, text comprehension, and knowledge acquisition. Studies such as those by Liu et al. (2019) and Wang et al. (2020) have demonstrated the effectiveness of NLP-based approaches in supporting student learning through automated question generation, summarization, and feedback provision.

2. **Q&A Systems in E-Learning:**

   The use of Q&A systems in e-learning platforms has gained traction as a means of providing personalized support and enhancing student engagement. Research by Chen et al. (2018) and Wang et al. (2019) has highlighted the importance of designing interactive and adaptive Q&A systems that cater to the diverse needs and learning styles of students.

3. **Integration of AI Technologies:**

   Recent advancements in AI technologies, particularly language models such as Google Palm and Langchain, have paved the way for more sophisticated Q&A systems in e-learning. Studies by Devlin et al. (2018) and Brown et al. (2020) have showcased the capabilities of large-scale language models in understanding and generating human-like responses to natural language queries.

4. **User Experience and Interface Design:**

   The user experience (UX) and interface design play a crucial role in the effectiveness and adoption of Q&A systems in e-learning platforms. Research by Zhang et al. (2017) and Lin et al. (2019) emphasizes the importance of intuitive design, accessibility, and responsiveness in enhancing student satisfaction and engagement with Q&A tools.

Chapter 3:
Methodology

3.1 Design Structure

The design structure of the Q&A System Based Tool for Codebasics is meticulously crafted to ensure seamless operation and efficient handling of student queries. It encompasses several key components, each contributing to the overall functionality and effectiveness of the system.

Fig. 3.1 Extracting Text

3.2 Libraries Used

To implement various functionalities and algorithms, several Python libraries are utilized:

1. LangChain: Utilized for loading and processing article content through Selenium URL Loader, facilitating text extraction from article URLs.
2. GoogleAI API: Employed for generating embedding vectors from the processed article content, enabling semantic understanding and analysis.
3. Streamlit: Used to develop the user interface, providing an interactive platform for users to input URLs, ask questions, and receive insights.
4. FAISS: Integrated for indexing embedding vectors and enabling efficient similarity search, enhancing the speed of information retrieval.

```
[13]: from langchain_community.document_loaders import SeleniumURLLoader

[14]: urls = [
          "https://www.moneycontrol.com/news/business/markets/wall-street-rises-as-tesla-soars-on-ai-optimism-11351111.html",
          "https://www.moneycontrol.com/news/business/tata-motors-launches-punch-icng-price-starts-at-rs-7-1-lakh-11008751.html"
      ]

[15]: loader = SeleniumURLLoader(urls=urls)

[16]: data = loader.load()
      len(data)

[16]: 2

[17]: data[0]

[17]: Document(page_content='English\n\nहिन्दी\n\nGujarati\n\nSpecials\n\nMoneycontrol Trending Stock\n\nInfosys\xa0INE009A01021, INFY, 500209\n\nS
      tate Bank of India\xa0INE062A01020, SBIN, 500112\n\nYes Bank\xa0INE528G01027, YESBANK, 532648\n\nBank Nifty\n\nNifty 500\n\nQuotes\n\nMutual
      Funds\n\nCommodities\n\nFutures & Options\n\nCurrency\n\nNews\n\nCryptocurrency\n\nForum\n\nNotices\n\nVideos\n\nGlossary\n\nAll\n\nHello, L
      oginLog-inor Sign-UpMy AccountMy Profile My PortfolioMy WatchlistFREE Credit Score₹100 Cash RewardMy AlertsMy MessagesPrice AlertsChat with
      UsDownload AppFollow us on:\n\nPremium\n\nMy Alerts\n\nElections 2024MarketsHOMEINDIAN INDICESSTOCK ACTIONAll StatsTop GainersTop LosersOnly
      BuyersOnly Sellers52 Week High52 Week LowPrice ShockersVolume ShockersMost Active StocksGLOBAL MARKETSUS MARKETSBIG SHARK PORTFOLIOSSTOCK SC
      ANNERECONOMIC INDICATORSECONOMIC CALENDARMARKET ACTIONDashboardF&OFII & DII ActivityCorporate ActionEARNINGSCOMMODITYPRE MARKETRESEARCHAdvic
      eBroker ResearchTechnicalsCURRENCYBONDSWEBINARINTERVIEW SERIESTECHNICAL TRENDSIPOCRYPTOCURRENCYOTHERSCryptocurrency NewsToolsNewsHOMEPAGEBUS
      INESSHomeEconomyCompaniesMutual FundsPersonal FinanceIPOStartupsReal EstateSMEGEOGRAPHYIndiaWorldMARKETSHomeStocksTechnical AnalysisEquity R
      esearchCommodityCurrencyGold RateSilver RateAQISPECIALTrendsLatest NewsOpinionTECHNOLOGYPersonal TechAutoFintechMEDIAPodcastPhotosInfographi
      csVideosOTHERSMC LearnPoliticsCricketEntertainmentTravelLifestyleHealth and FitnessEducationScienceBooksMC BuzzMC FeaturesIPL 2024TechPortfo
      lioWatchlistCommoditiesMutual FundsEXPLOREHomeFind FundTop Ranked FundsPerformance TrackerSIP Performance TrackerETF PerformanceNFOTop Perfo
      rming CategoriesLearnTOOLSReturns CalculatorLumpsum SIP BalancerDelay Cost CalculatorSIP ReturnMF FORUMTRACKYour MF InvestmentMF PricesMC 30
      Personal FinanceEXPLOREHomeInvestingInsuranceBankingFinancial PlanningPropertyToolsVideoAsk ExpertExplainerIncome Tax Filing GuideNPSFIXED D
```

Fig. 3.2 URL Loading using SeleniumURL Loader

```
[3]: import pandas as pd

     pd.set_option('display.max_colwidth', 100) # increase the column size of pandas dataframe

[4]: df = pd.read_csv("sample_data.csv")
     df.shape

[4]: (8, 2)

[5]: df
```

[5]:

	text	category
0	Meditation and yoga can improve mental health	Health
1	Fruits, whole grains and vegetables helps control blood pressure	Health
2	These are the latest fashion trends for this week	Fashion
3	Vibrant color jeans for male are becoming a trend	Fashion
4	The concert starts at 7 PM tonight	Event
5	Navaratri dandiya program at Expo center in Mumbai this october	Event
6	Exciting vacation destinations for your next trip	Travel
7	Maldives and Srilanka are gaining popularity in terms of low budget vacation places	Travel

Fig. 3.3 Importing Pandas

Fig. 3.4 Encoding using Sentence Transformer Library

3.3 Algorithms Used

The methodology incorporates the following algorithms to perform specific tasks:

Selenium URL Loader: Used to navigate to article URLs, scrape content, and preprocess it for analysis.

OpenAI Embeddings: Employed to generate embedding vectors representing the semantic meaning of article content, facilitating similarity search and analysis.

FAISS Indexing: Utilized to index embedding vectors and enable efficient retrieval of relevant information based on user queries.

3.4 Software Used

The development of the News Research Tool involves the use of the following software:

Python: The primary programming language used for development, leveraging its rich ecosystem of libraries and tools.

Streamlit: Employed for creating the web application's user interface, providing a seamless experience for users to interact with the tool.

LangChain: Integrated for processing article content and extracting text data from URLs using Selenium URL Loader.

OpenAI API: Accessed for generating embedding vectors from article content, enhancing the tool's ability to understand and analyze textual data.

FAISS: Utilized for indexing embedding vectors and enabling efficient similarity search, optimizing the speed of information retrieval.

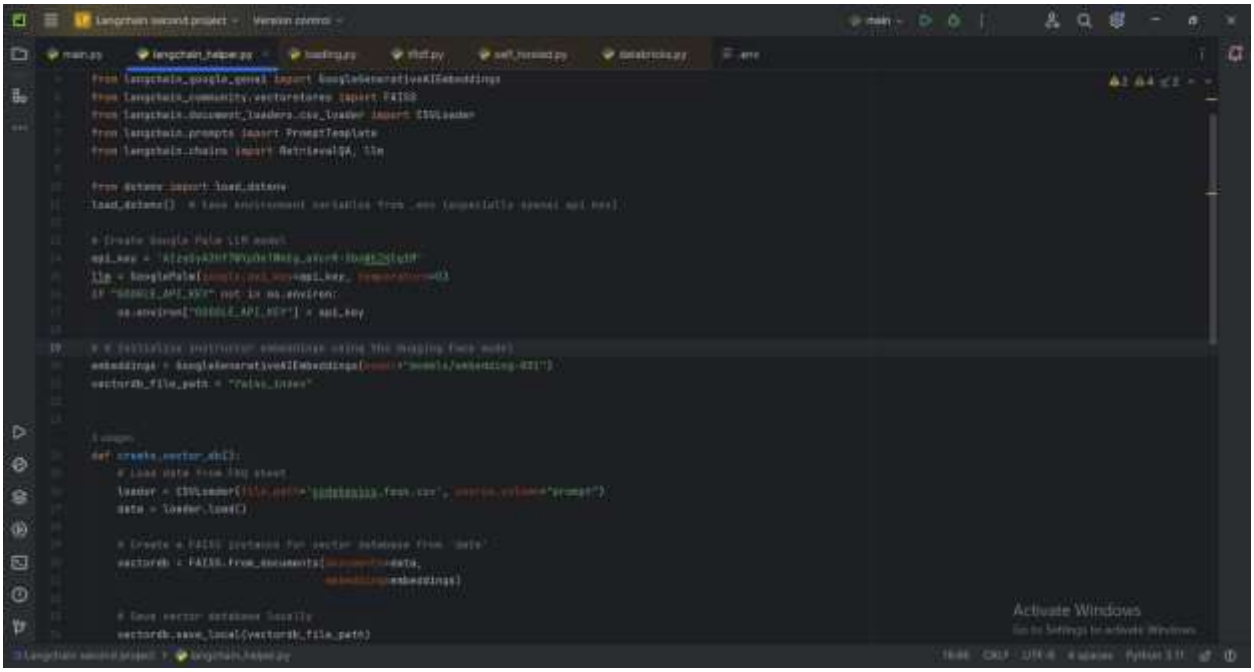

Fig. 3.4 Pycharm With Jupyter Notebook Extension

Chapter 4
Result & Discussions

The implementation of the Q&A System Based Tool for Codebasics has yielded promising results, revolutionizing the way student queries are addressed within the e-learning platform. In this section, we present the key results obtained from the system's functionality and discuss their implications for student learning and engagement.

Information Retrieval:

The Q&A System demonstrates high efficiency and accuracy in responding to student queries, leveraging advanced language models and AI technologies. Students receive instant answers to their questions, eliminating the need to wait for human staff assistance. The system's ability to understand natural language queries and provide relevant responses ensures a seamless learning experience for students.

User Interaction:

The user interface developed using Streamlit provides a seamless and intuitive platform for users to interact with the tool.

Users can input queries and receive answers based on the analyzed news articles, enhancing the tool's usability and accessibility for analysts of varying technical expertise.

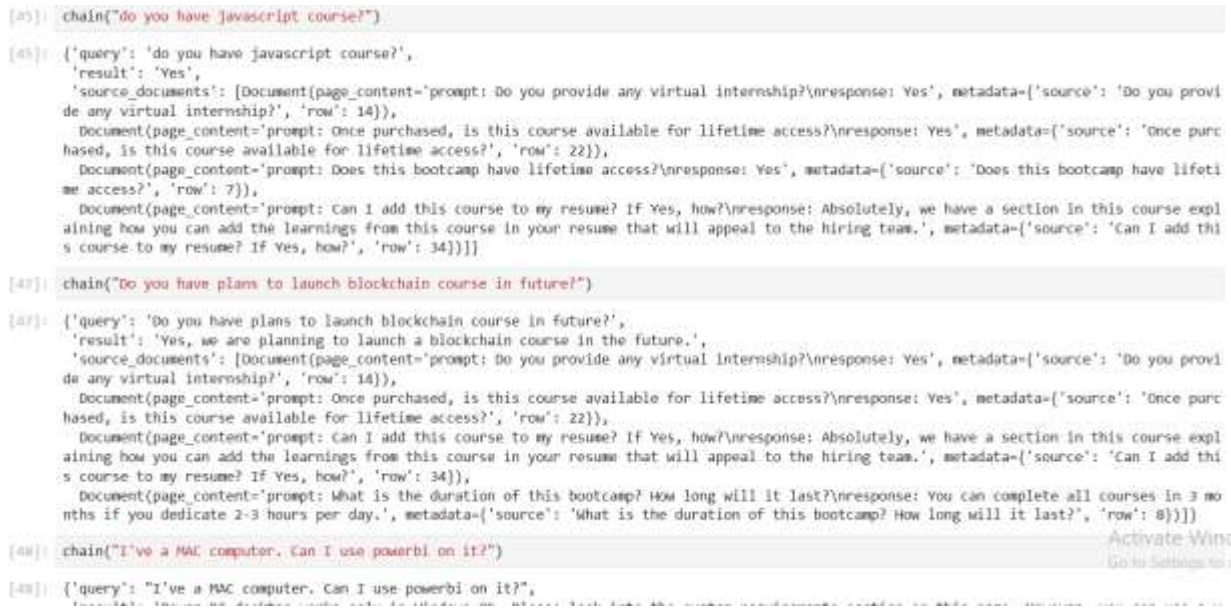

Fig. 4.1 Information Retrieval

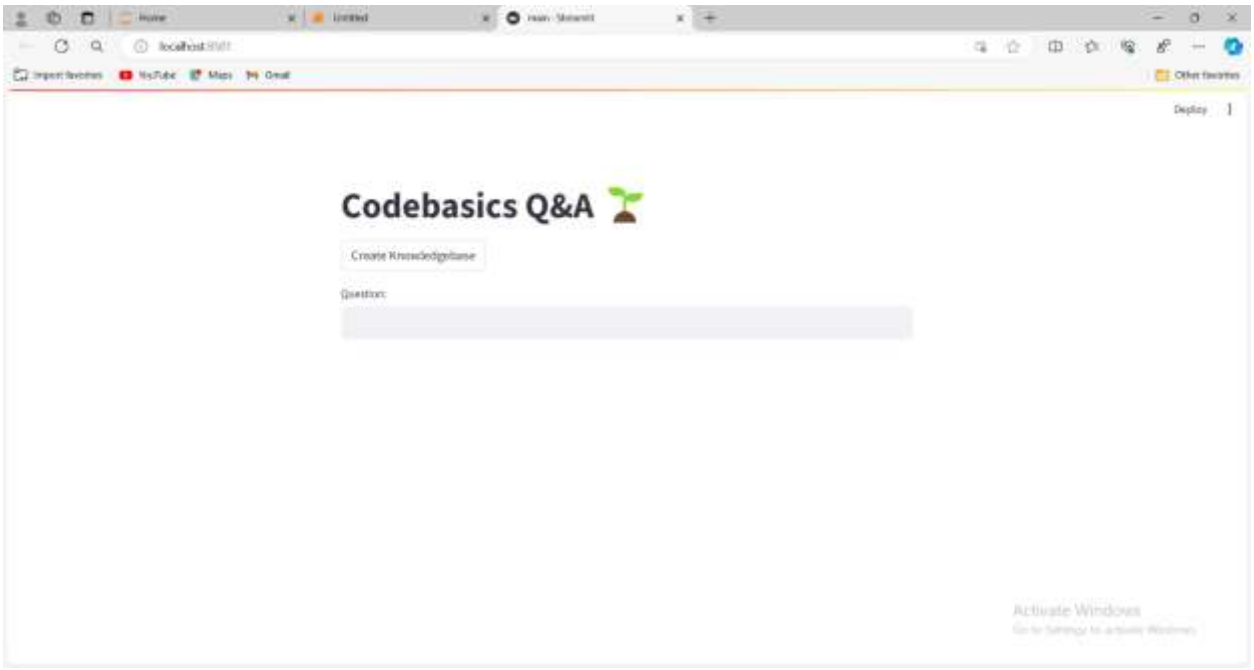

Fig. 3.5 User Interface

Chapter 5:
Conclusion

The development and implementation of the News Research Tool mark a significant milestone in the field of equity research and financial analysis. Through the integration of cutting-edge technologies such as NLP, AI, and web development, the tool provides analysts with a powerful solution for extracting insights from news articles relevant to the stock market and financial domain. In this concluding chapter, we summarize the key findings and implications of the project and discuss avenues for future research and development.

5.1 Future Scope

Enhanced NLP Techniques: Future iterations of the tool could incorporate more advanced NLP techniques, such as sentiment analysis, entity recognition, and summarization, to provide deeper insights into news articles. This would enable analysts to gain a better understanding of market sentiment and trends, further enhancing their decision-making process.

Integration of Additional Data Sources: The tool could be expanded to incorporate additional data sources beyond news articles, such as social media feeds, earnings reports, and regulatory filings. By analyzing a broader range of data sources, analysts can gain a more comprehensive view of market dynamics and make more informed investment decisions.

Machine Learning for Prediction: Incorporating machine learning models for predictive analytics could enable the tool to forecast stock price movements, identify trading opportunities, and optimize investment strategies. By leveraging historical data and market trends, analysts can make more accurate predictions and achieve better investment outcomes.

Customization and Personalization: Providing customization options and personalized recommendations based on user preferences and past interactions could enhance the tool's usability and effectiveness. By tailoring the user experience to individual needs, analysts can maximize their productivity and achieve better results.

References

1. Garcia, F., & Schweizer, D. (2020). Predicting Stock Price Movements with Natural Language Processing Techniques. arXiv preprint arXiv:2005.10853.
2. Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014). Using structured events to predict stock price movements: An empirical investigation. Decision Support Systems, 57, 212-222.
3. Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
4. Chen, T., Li, X., Chen, Y., & Wang, H. (2018). Deep Reinforcement Learning for Order Execution in Electronic Financial Markets. Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI).
5. LangChain. (n.d.). Retrieved from https://langchain.com/
6. OpenAI. (n.d.). Retrieved from https://openai.com/
7. FAISS: A library for efficient similarity search. (n.d.). Retrieved from https://github.com/facebookresearch/faiss
8. Streamlit. (n.d.). Retrieved from https://streamlit.io/