Shubham Kumar

Email: shubhamkr2091@gmail.com
Mobile: +91-620-4693-362

Portfolio: https://shubham2091.vercel.app
Github: github.com/zyberg2091

EDUCATION

- **SRM Institute of Science and Technology**   Chennai, India
  Bachelor of Technology - Computer Science; GPA: 9.12   *June 2018 - June 2022*
  Courses: *Data Structures, Analysis Of Algorithms, Artificial Intelligence, Machine Learning, Databases, Probability and Statistics*

SKILLS SUMMARY

- **Languages**:   Python, C++, SQL
- **Frameworks**:   Scikit, NLTK, SpaCy, TensorFlow, Keras, Flask, Pandas, RASA
- **Tools**:   GIT, Heroku, MySQL
- **Platforms**:   Web, Windows, Google Colab, Anaconda, VSCode, AWS (Familiar with)
- **Soft Skills**:   Leadership, Cross Functional Team Building, Team Work

EXPERIENCE

- **HighRadius**   Remote
  Summer Intern   *June 2021 - Present*

  B2B Invoice Management System

  - **ML Model API**: performed data preprocessing, EDA(exploratory data analysis), feature engineering (built multiple new features), trained a **gradient boosting** machine learning model on 40k size of Invoice data and exported the predicted CSV file of size 10k along with ML model.
  - **Database Integration**: Converted csv file to sql file using **sqlalchemy** and stored the data into database table of mysql server using an automated python script.This data is fetched in the backend by creating a connection between backend and sql server using JDBC driver and displayed on the UI of Dashboard.

- **Pain Free Academy**   Remote
  Data Science Intern (Part-time)   *Sep 2020 - Dec 2020*

  - **Project**: Built machine learning/deep learning models using a variety of frameworks such as tensorflow and scikit learn leveraging NLP in the verticals of healthcare and smart city analytics.
  - **Company Blogs**: I handled the responsibility for creating AI/ML contents for the company's blog website.

PROJECTS

- **Healthify bot (NLP, Chatbot, Interactive)**: A Telegram bot that solves your daily medical queries using SOTA ML algorithms built with Rasa. It is based on a **multi-task transformer architecture (DIET)** which is **6 times faster** to train than **BERT**.

- **Aapka Apna Hip-Hop (NLP, Web)**: The next generation Rap tool which uses a sequence language model **LSTM** to generate lyrics for your own rap songs lyrics in the style of a specific artist **supported by 3 artists**, i.e. Eminem, Drake Kanye West. Their respective song data was collected through web scraping and **EDA** was performed on those data.

- **BlogTagger (Tags, Python Package (pre-released))**: This Package can be used to extract keywords from a page to create tags for any blogs,news or any textual information available on the web page.sentence embedding vectors of tokens and Text (collection of tokens) were extracted using state of the art **ALBERT** model (small size variant of BERT model).

- **Sehpathi (Web, Collection of best youtube courses)** : currently in development phase - A website with the motive of allocating best courses for any new aspiring developers. **Multilple roadmaps** on different stacks has been created for the developers based on youtube lectures playlist extracted using **youtube API**.

RESEARCH PAPERS

- **Smart system to detect adult content and child pornography on web (July'21)**: An architecture designed to predict toxicity of web pages, adult video embedded in web pages.It automatically blocks the adult video if any content in the video represents child pornography.

HONORS AND AWARDS

- 3rd prize in HackCBS 3.0 (Major League Hackathon) - Nov, 2020
- 3rd prize at HackYuva 2020 (MyWays and LearnYuva) - Oct, 2020
- Top 15 in SLAC 2020 (Amrita School of Engineering,Bangalore) - Nov, 2020
- Top 15 in HackThisFall 2020 - Oct, 2020

ORGANIZATIONS

- **AI Research Associate at Next tech Lab,SRM Chennai**   Mar 2020 - April 2021
  Next Tech is a multi-disciplinary lab founded by MIT alumni,Silver Winner of QS re imagine education.