SHLOK VIVEK NAIK

Raleigh, North Carolina | naikshlok.sn@gmail.com | www.github.com/shlok00 | www.linkedin.com/in/shlok-naik/

EDUCATION

North Carolina State University, Raleigh, NC - Master of Computer Science **Aug 2022 - Dec 2023**

Relevant Coursework: Automated Learning and Data Analysis, Design and Analysis of Algorithms, Neural Networks and Deep Learning, Automated Software Engineering, Foundations of Data Science

SIES Graduate School of Technology, University of Mumbai, India - Bachelor of Computer Engineering **Aug 2018 - Jun 2022**

Relevant Coursework: Machine Learning, Database Management Systems, Data Warehousing and Mining, Data Structures, Big Data Analytics, and Natural Language Processing.

TECHNICAL SKILLS

- **Languages**: Python, C/C++, SQL, JavaScript, HTML/CSS, R, Kotlin
- **Frameworks and Libraries**: Flask, Bootstrap, NodeJS, MongoDB, Express.js, React.js, Next.js, pandas, NumPy, Matplotlib, Sklearn, TensorFlow, OpenCV, NLTK, SpaCy, torch, Gensim, PHP, AJAX, JQuery
- **Developer Tools**: Github, Heroku, Google Cloud Platform, Android Studio, PyCharm, Spyder, Jupyter Notebooks, Eclipse, Atom, Google Colab, Amazon Web Services, Google Cloud Platform, Tableau

PROJECTS AND PAPERS

Assignment Submission and Marking Portal

- Built an efficient assignment checking and submission interface using **React.js, Node.js,** and **Express.js**, with an ability to check for plagiarism and cheating in a timely manner averaging 5 seconds per assignment.
- Developed a sophisticated plagiarism detection system that factors in the sentence structure and context of two paragraphs. It achieved an 85% accuracy rate. Additionally, integrated Gensim's **Doc2Vec** model and **SpaCy** and framed an API endpoint for easy access.
- Demonstrated proficiency in object detection techniques utilizing **p5.js, YOLO**, and **OpenCV** frameworks to track and analyze eye movement patterns of a test subject, detecting any unusual or suspicious activity. Coordinated the integration of subject and additional human detection functionality into video frames. Deployed the website on Heroku and implemented Cloudinary for efficient data storage.

Cash Flow Management Application

- Developed a mobile application that extracts information from bills and generates an organized financial statement for the user. Utilized Flutter and **DART** to create a user-friendly interface that accepts images of bills.
- Implemented image segmentation with **OpenCV** that divides the bill into text and whitespace blocks. Integrated Optical Character Recognition (**OCR**) to extract text and preprocess it to create a dataset. Trained a Sequential Model to classify bill elements into different categories based on the dataset. During application testing, achieved an accuracy rate of 75% and a processing time of 10 seconds per bill.

SponsEnd - Sponsorship Advertisement Filtering Tool

- Formulated a web plugin that detects sponsored content in YouTube videos by taking in a URL link as input. Utilized the YouTube transcripts API to retrieve video text and timestamps, preprocessing it for classification tasks. Employed **Doc2Vec** model and **MLPRegressor** Autoencoder for classification on SponsorBlock's open-source dataset. Created a customized confidence index to determine the probability of a tagged segment being sponsored content. Hosted the site on **Amazon EC2** for efficient accessibility.

Smart Journal Application for Mental Health Analysis

- Developed a mobile application using **Kotlin** and **Firebase** to create a sophisticated mental health monitoring platform for individuals by analyzing hundreds of textual entries and generating insightful visualizations.
- Implemented voice-to-text functionality in multiple languages, enabling users to easily document their daily experiences. Employed artificial neural networks (**ANN**) in **Python** to develop Mood Detection and Profanity Detection features, achieving high accuracy rates of 80% and 85%, respectively.

Employee Attrition Prediction and Retention Strategy

- Pioneered the development of a predictive tool utilizing **SVM, XGBoost, and Naive Bayes** algorithms augmented by Principal Component Analysis. Used Python and scikit-learn to determine employee attrition risk rates in healthcare.
- Spearheaded a comparative analysis to determine the most effective method, employing analytical tools such as **KNN** to identify the top 10 at-risk employees and utilizing comparative methods to discover optimal parameters for employee retention.
- Achieved a 92% accuracy rate using the XGBoost algorithm, demonstrating exceptional precision and accuracy in predicting employee attrition risk rates. Hosted the application using a Flask server, ensuring seamless and efficient deployment.

Application for neutralizing socially sensitive content in the webspace

- Developed a tool for detection and neutralization of text on the internet which qualifies as sexist, racist or queer phobic. Collated and processed data from over 20 datasets. Implemented a chrome-based plugin for the same using **Manifest v-2** and hosted the tool on this.
- Utilized an **XLNet** and **RoBERTa** model for detection of sensitive keywords and utilized the **PEGASUS** model for neutralization. The model provided an accuracy of around 75%.

PROFESSIONAL EXPERIENCE

SIES Graduate School of Technology, Mumbai, India - *Research Intern* **Jun 2021 – Aug 2022**

- *Contributed to the publication of the research paper 'Investigating Clinical Named Entity Recognition Approaches for Information Extraction from EMR.* Led the development of a treatment prediction model for diseases by extracting relevant symptoms and tests from over 100,000 lines of textual data scraped from Electronic Medical Records of 50 patients using SpaCy. Categorized the extracted data points according to the BILUO notation to enable efficient processing and analysis.
- Developed a named entity recognition system that categorized entities in Electronic Medical Records (EMRs) using Spacy and Google's BioBERT model, achieving an accuracy rate of 80%. Created visual representations of 100 identified diseases, symptoms, and treatments using Neo4.js and saved them to a Neo4.js database for efficient EMR entity linking which facilitated disease and treatment prediction based on symptoms. Significantly reduced manual processing time by 85%.

Widhya, Mumbai, India - *Data Science Intern* **Jan 2021 – Feb 2021**

- Accomplished multiple data analysis projects, including Covid-19 analysis, flight delay prediction, stock price prediction and Instagram post reach prediction with the guidance of a mentor. Devised prediction engines and visualization dashboards using seaborn. Demonstrated expertise in Python by skillfully integrating diverse concepts such as data visualization, time series analysis, data preparation, and model testing.

SmartBridge Educational Services Private Ltd., Mumbai, India - *Machine Learning Intern* **May 2020 – Jul 2020**

- Developed an Intelligent Customer Helpdesk for Water Purifiers by automating query resolution via application of Smart Document Understanding function on Product Manuals. Leveraged IBM Watson Assistant, Discovery and IBM modules to train the Helpdesk chatbot using the QnA bank generated by the function. This consequently reduced human intervention by 90%.

AWARDS & ACHIEVEMENTS

- Technical Excellence Award winner - SIES GST Batch of 2022 for outstanding performance in technical domain among 450 students.