SAI HARSHA VARDHAN REDDY KOLAN

Hayward, CA 94542 | +1(925)-523-8123 | skolan@horizon.csueastbay.edu

LinkedIn   Portfolio

Education

California state University, East Bay	**Jan 2023-Dec 2024(expected)**
Master of Science in Business Analytics:	**CGPA: 3.628/4**

Relevant Coursework: *BAN 610*-Database Management and Applications, *BAN 612*-Data Analytics, *BAN 620*-Data Mining, *BAN 622*-Data Warehousing and Business Intelligence, *BAN 630*-Optimization Methods for Analytics, *BAN 632*-Big Data Technology and Applications, *BAN 674*-Machine Learning for Business Analytics, *BAN 673*-Time Series Analytics, *BAN 675*-Text Mining and Social Media Analytics

Sreenidhi Institute of Science and Technology	**Jun 2018-Jun 2022**
Bachelor of Technology in Electronics and Communication Engineering:	**CGPA: 3.5/4**

Skills

Problem-Solving:	Python Hackerrank Certification, SQL Hackerrank Certification.
Programming Languages:	Python [statsmodels, Numpy, Pandas, Matplotlib, Seaborn, BeautifulSoup, Scikit-learn, XG-boost, Keras TensorFLow, NLTK, spaCy], SQL, R, JavaScript, C++.
Statistics:	Hypothesis testing [z-tests, t-tests, f-tests, chi-squared tests, ANOVA], Regression Analysis, A/B testing.
Software:	IDE [Jupyter Notebooks, PyCharm], Databases [MySQL, Microsoft SQL Server], Spreadsheets [MS-Excel].
Data Visualization:	Tableau, Power BI.
Big-Data:	Apache Hadoop, PySpark, Apache Hive, Apache Pig.
Cloud:	Azure Studio, Azure Databricks, AWS-EMR, AWS-RedShift, AWS-Glue Studio, Google-Big Query.
Predictive Modeling:	Linear Regression, Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, Naive Bayes, SVM.

Work Experience

AIVID TechVision	**Ahmedabad, India**
Machine-Learning intern	**Oct 2022-Dec 2022**

- Annotated on an average of 1000 frames per day using V7-Darwin annotation tool, making a substantial 60% contribution to the training data used in the development of face detection and security-surveillance AI Bots.
- Collaborated with senior machine learning engineers, contributing to data cleaning, preprocessing, and fine-tuning Faster R-CNN models to enhance object detection efficiency. This collaborative effort resulted in a notable 5% increase in model accuracy.

Projects

BlackCoffer's Web-Articles Analysis Using Text-Mining and NLP	**Jan 2024-Mar 2024**

- Conducted web scraping of 89 diverse text-articles and implemented text preprocessing techniques including tokenization and stop-word removal to prepare data for analysis, demonstrating proficiency in data acquisition and preprocessing.
- Executed sentiment analysis on the collected web articles, computing polarity and subjectivity scores to discern the underlying sentiments, showcasing strong skills in natural language processing (NLP) and text analysis.
- Utilized NLP methodologies like TF-IDF and chunking to extract relevant information from both highly-ranked and lower-ranked articles, demonstrating proficiency in employing NLP for effective information retrieval.

Big-Data Applications in USAF Weather Station Analytics	**Oct 2023-Dec 2023**

- Developed a MapReduce application to efficiently calculate monthly averages of wind direction for each year in a decade, showcasing expertise in parallel computing and large-scale data processing.
- Developed a PySpark application to compute the range of sky ceiling height for each weather station ID, demonstrating proficiency in Python.
- Utilized Hive and Pig for calculating average visibility distance and range for each USAF weather station ID, demonstrating adeptness in SQL for data querying and processing within distributed environments.

DDos-detection using Machine-Learning	**Sep 2023-Oct 2023**

- Developed a real-time random forest machine-learning model using Azure machine learning studio cloud service for efficient detection of distributed denial of service attack in application layer of networking model.
- The model is trained with the dataset having 7.6M datapoints (1.3M DDoS, 6.3M Benign) using hyperparameter tuning by implementing grid-search with 3-fold cross-validation technique and executed on Standard_E8s_v3 (8 cores, 64 GB RAM, 128 GB disk) compute instance.
- The model has an impressive accuracy of 99.9% with only 4 misclassifications on validation data along with excellent values of precision-1.0, recall-0.999, F1-score-0.999 indicating the model will perform exceptionally well for DDoS detection.

Cricket-Analytics (ICC-T20 World Cup-2022)	**July 2023-Aug 2023**

- Conducted web scraping of Cricket T20 World Cup 2022 data from the official ESPN website using JavaScript.
- Utilized Python and the Pandas library to parse JSON files, preprocess, clean, and transform the retrieved data, ensuring high data quality.
- Performed data modelling, formulated around 20 DAX measures and designed an interactive dashboard in Power BI to present insights into top power hitters, middle-order batsmen, all-rounders, and leading bowlers, showcasing data visualization and storytelling capabilities.

Handoff using Machine-Learning techniques	**Nov 2021-Mar 2022**

- Performed an extensive literature review to bridge the gap between data analysis and handoff procedures in telecommunication networks, identified key-factors in handoff process between central nodes and mobile devices for seamless connectivity.
- Generated a real-time dataset and pre-processed it using python to facilitate the efficient development of machine learning models for handoff identification.
- Developed logistic regression, decision trees, and random forest models to identify handoff occurrences, subsequently analyzing their accuracies and confusion matrices. Random forests exhibited the highest accuracy of 87% in predicting handoffs.