Lokesh Kank

San Diego, CA, USA | +18587442548 | LinkedIn | kank.lokesh75@gmail.com | GitHub | Portfolio

EDUCATION AND TRAINING

San Diego State University (SDSU) Aug 2022 - Aug 2024

Master of Science in Big Data Analytics (BDA) / Data Science San Diego, CA, USA

Relevant Coursework: *Big Data Science and Analytics Platforms, GIS Programming with Python, Business Analytics, Machine learning Engineering, Scientific Database management.*

Savitribai Phule University Aug 2016 - May 2019

Bachelor in Mechanical Engineering – GPA 3.4/4 Pune, MH, India

SKILLS AND KNOWLEDGE

Programming skills & technology: Python, R, SQL, Postgres, HTML, CSS.

Data Science, AI & Machine Learning: Pandas, Numpy, Scipy, Plotly, Pyspark, Matplotlib, Pytorch, Tensorflow, Scikit-learn, NLTK, Natural Language Processing (NLP), ETL, Model Validation, Statistical Modeling.

Toolkit: AWS, Snowflake, Tableau, Databases (Mysql, MySQL), Github, Jupyter, Pycharm, Microsoft Excel

Soft Skills: Verbal and Written Communication, Self-motivated, Innovative, Analytical, and Problem-Solving.

PROFESSIONAL EXPERIENCE

Data Science Intern (Data Glacier) Feb 2023 – Present

- Working on a project to develop a model to predict whether a customer will buy Bank's term deposit product. Using agile methodologies such as **Scrum** & **Kanban** to plan and track the progress.
- Conducting exploratory data analysis and data preparation tasks. Using **Bitbucket** for version control and code review to ensure code quality and maintainability.
- Building various machine learning models, including logistic regression, ensemble, and boosting, to predict customer behavior and evaluate their performance. Developing **Flask** web applications and deploying applications to the **cloud** and integrating with **APIs** for seamless data retrieval and processing.

Graduate Research Assistant (Department of Computer Science, SDSU) Jan 2023 – Present

- Working on Anomaly Detection of microvascular RBC flow and velocity analysis with pupil tracking/timing.
- Researched the use of **CNN** models using data in .avi video format. Utilized Python, subprocess, & FFmpeg modules and converted ".avi" to "mp4" format, allowing for easier processing.
- Leveraged the **OpenCV2** video capture class and extracted metadata and images from videos to create datasets. In collaboration, design a future approach for the project using the IDEAL mechanism.

Software Engineer (Accenture, Advance Technology Centre) Feb 2020 – Jul 2022

- Exploited SQL, and Python for failure analysis**,** following best practices for fetching business insights for reporting and best user experience. Saved **80** hrs of work by collaborating to develop automation in python.
- Experienced in migrating applications on Linux machines & work with cross-functional teams with a vision.
- Deploying hotfixes in a development phase, and releases in the PROD environment. The release has contributed **3.2** % **($ 7.2 Million**) of quarterly profit and increased its sales growth.

RELEVANT PROJECTS

Prevention of medication error using Deep Learning (https://sites.google.com/sdsu.edu/)

- Analysed the problem of Medication errors on various drugs and developed DL models (neural network) – VGG 16, Xception, and Inception with **95 to 99%** accuracy to classify pharmaceutical drugs efficiently.
- Implemented Image Preprocessing (Edge detection, Masking, standardization, and normalization) to raise the image's quality and reduce medication errors by **56%**.
- Integrated **Tableau** interactive dashboard for business intelligence about medication errors in the medical field.

Infer energy star scores of new buildings

- Conducted data extraction, Data preprocessing, and Built regression/classification models that can estimate a building's Energy Star Score based on selected features after feature engineering (one hot encoding, Removed collinear features with collinear coefficient > **0.6**).
- Established baseline error scores, and compare different models. Used Random Forest with **MEA 9.044**.

Time series forecasting of energy generation data

- Reduced loss due to delays in the production system by **68%** by forecasting undesirable breakdowns.
- Increased energy production by **47%** by analyzing past energy generation data of wind turbine generators.
- Fitted, evaluated, & made predictions with the Random Forest model for time series forecasting.
- Created a Powerbi Dashboard to display the forecasted energy and breakdown data.

Certification

AWS - Certified Solution Architect Issue on – May 2022 – Expire on – July 2025