Lokesh Kank

San Diego, CA, USA | +1 (858)-744-2548 | LinkedIn | kank.lokesh75@gmail.com | GitHub | Portfolio

EDUCATION

San Diego State University (SDSU) Aug 2022 - May 2024
Master of Science in Big Data Analytics (STEM) San Diego, CA, USA
- *Big Data Science and Analytics Platforms.* - *GIS Programming with Python.* - *Business Analytics.*
- *Database Theory and Implementation.* - *Machine learning Engineering.* - *Data Management for GIS.*

Savitribai Phule Pune University Aug 2016 - May 2019
Bachelor in Mechanical Engineering – GPA 3.4/4 Pune, MH, India

SKILLS

Programming & Technology: Python, R, SQL, PostgreSQL, Matlab, CSS.

Machine Learning Techniques: Pandas, Numpy, Pyspark, Keras, Matplotlib, Pytorch, Tensorflow, Scikit-learn algorithms, Data Visualization, Sparks, Exploratory Data Analysis, Statistics, NLP, Hadoop, Applied Mathematics.

Toolkit and Knowledge: AWS Sagemaker, Databricks, Git, Docker, MongoDB, Hive, Github, Pycharm, Excel.

EXPERIENCE

Data Science Intern Feb 2023 – Present
Data Glacier San Diego, CA, USA
- Developing a predictive model for determining the likelihood of customers buying the Bank's term deposit product. Reduced **37%** of defects and increased productivity by **16%** as measured by defect density and burn-down chart after implementing Agile methodologies (Scrum and Kanban).
- Attained a **7%** increase in model accuracy by handling missing values, outliers, and performing scaling and normalization to reduce predictor biases. Utilized correlation matrix for getting insights, resulting in a model accuracy of over **92%**.
- Achieved a **50%** increase in stakeholder trust & reliability measures by effectively communicating project progress and statistical insights, which allowed for a reduction in meeting frequency from twice a week to once a week.

Software Engineer Feb 2020 – Jul 2022
Accenture, Advance Technology Centre Pune, MH, India
- Obtained a **35%** increase in data retrieval speed, by streamlining SQL queries and using indexing strategies.
- Exploited SQL, and Python for troubleshooting, and created business reports for client presentations. This resulted in a **26%** increase in client satisfaction as measured by NPS (Net Promoter Score).
- Saved **60hrs** of monthly work by developing automation in python resulting in smooth processing of cross-functional.
- Concluded **18%** improvement in the application performance as gauged by the Smile tool after migrating the application to a new RHEL Linux server

PROJECTS

Prediction of energy star scores of new buildings
- Accomplished a **12%** increase in model accuracy examined by error reduction and by conducting data processing, EDA, and feature engineering, including one hot encoding and removal of collinear features with a collinear coefficient > **0.6**. Additionally, selected essential predictors, taking note of negatively correlated predictors.
- Established baseline error of **25%** by taking a median of the training label set which helped compare ML models.
- Performed hyperparameter tuning on the best model causing an increase in model accuracy by **8%.** Evaluated the best model on the testing set and hence went with Random Forest with MAE **9.044**.

Classification of pharmaceutical drugs for prevention of medication error (website)
- Reduced medication errors by **56%** after implementing Image Pre-processing to raise the image's quality.
- Analysed the problem of Medication errors on various drugs and developed Deep Learning (Neural Networks) models – VGG 16, Xception, and Inception with **95 to 99%** accuracy to classify drugs efficiently.
- Integrated Tableau interactive dashboard for business intelligence about errors in the medical field and created a project overview using Google sites.

Data science salary recommendations
- Scrapped about **1000** job listings from Glassdoor using Selenium for datasets by understanding data architecture.
- Cleaned **80%** of scraped data and turned categorical data into numerical labels using label encoding for model building
- Split the data into train & test sets with **20%** for testing to train the model and evaluate its performance.
- Achieved MAE of **11.22** by using the Random Forest model. Used linear regression as a baseline model.

CERTIFICATION

AWS - Certified Solution Architect May 2022 – Jul 2025