SANKETKUMAR PATEL

+1(312) 647-6160 ⋄ Chicago, IL (Willing to Relocate)
spatel194@hawk.iit.edu ⋄ linkedin.com/in/sanketkumarpatel07/ ⋄ https://github.com/SanketKumarP

OBJECTIVE

A detail-oriented data analyst with a passion for uncovering insights through data analysis and developing accurate predictive models. Skilled in utilizing statistical and machine learning techniques to drive decision-making and enhance business performance. Seeking to leverage expertise in a data analyst role to contribute to the organization.

EDUCATION

Master of Data Science, Illinois Institute of Technology — Enrolled: Jan 2023 — Expected: Dec 2024
Relevant Coursework: GPA 3.83
Applied statistics, Monte-Carlo methods in Finance, Statistical learning, Time series, Probability and Statistics, Data Preparation and Analysis, Big DATA, Machine learning, Database Organization, Data Science Practicum, Public engagement for Data Scientists

SKILLS

Programming:	Python, R, MySQL, MATLAB, Java
Frameworks:	Hadoop, Spark, Kafka, Hive
Libraries:	NumPy, SciPy, Matplotlib, MLlib, Scikit-learn, PySpark, dplyr, ggplot2, caret, itsmr
Cloud Platform & Tools:	Amazon Web Services (AWS), Tableau, PowerBI, Excel, GitHub
Certificates	Python for Data Science, AI & Development By IBM

PROJECTS

Exploring Performance and Efficiency in the CTA System: A Data-Driven and Geospatial Analysis.

- Conducted analysis of **Chicago Transit Authority (CTA)** system, identified busiest routes and stations, analyzed ridership load changes on different stations, and detected service delays.
- Preprocessed and cleaned datasets with imputation and outlier detection techniques utilizing concepts of **applied statistics**, ensuring data consistency in **Rstudio**.
- Utilized **regression models**, including simple linear and polynomial regression, to evaluate ridership trends and identify peak days, informing strategies for system efficiency.
- Through **API**, utilized real-time data to Develop a data-driven approach to identify average travel wait times.
- Deployed geospatial analysis using **Leaflet package** in **R**, providing user-friendly visualizations of station locations and ridership data.
- Extracted insights on High Ridership in Loop, Airport, and other stations, Impact of COVID-19 on Ridership, Post-COVID Recovery, and Seasonal Variation in Ridership.

Big data Technologies and engineering with aws and Hadoop

- **AWS Cloud and Big Data Infrastructure**: Set up and managed **AWS** services including **EMR** (Elastic MapReduce), **S3** (Simple Storage Service), and **Kafka** for data processing and streaming applications.
- **Hadoop Cluster Administration**: Created and managed **Hadoop** clusters on **AWS EMR**, including configuration and deployment of jobs using **HDFS** (Hadoop Distributed File System).
- **Kafka Data Streaming**: Implemented **Kafka** producers and consumers on AWS EMR for real-time data processing, ensuring efficient data ingestion and distribution across applications.
- **Data Processing with Hive**: Developed **Hive** queries and programs to analyze and process large datasets stored in **Hadoop**, leveraging Hive's SQL-like interface for data manipulation.

Airline Passenger Demand Forecasting using Time series.

- Developed a predictive model in **Python** for various **Time Series models** such as **ARMA, ARIMA, SARIMA, and Prophet** unsing python libraries **sklearn, statsmodel pmdarima** for accurate airline passenger demand forecasting, optimizing resource utilization, revenue, and cost management for airlines.
- Achieved 93% accuracy with best performing model **SARIMA**.

Bankruptcy Prediction using Machine Learning and financial reports in python .

- Developed Bankruptcy prediction models in **Python** using **matplotlib, sklearn, and statsmodels** libraries, addressing data imbalance with **SMOTE** technique, **normalizing** data, and resolving collinearity with **PCA**. Employed Recursive Feature Elimination (**RFE**) algorithm to eliminate features that are not helping in prediction of target variables.
- Developed 11 different **Machine learning models**, and concluded **XGBoost** as top classifier for accurate predictions with an accuracy of 99.3%.

EXPERIENCE

Graduate Engineer June 2019 - Nov 2022
Isgec Hitachi Zosen Ltd *Gujarat, India*

- Leveraged **SQL** to manage and view data from the **SAP**, Analyzed and optimized progress in 20+ projects, enhancing manufacturing and material procurement efficiency; presented findings to company officials.
- Implemented **Python** programs for mechanical strength calculations, reducing manual effort by 70%.
- Developed a **MATLAB** program for crack detection and weld length measurement, achieving an 80% efficiency.
- Managed and Designed projects for critical equipment, including Ammonia converted shell, CCR reactor, and others, ensuring timely delivery.

Co-founder & Lead of Analysis and Design Team Feb 2016 - Jan 2019
Team Sovereign *Gujarat, India*

- Co-founded team Sovereign in university to build an Electric all-terrain Vehicle (E-ATV) to participate in various competitions.
- Developed the design of E-ATV.
- Managed team funds, and secured sponsorships worth $20k, overseeing material procurement and fabrication outsourcing.