# Vatsal Jigar Kadakia

Hoboken, NJ (813) 843-9863 vatsalkadakia20@gmail.com LinkedIn Portfolio

## Education

**Stevens Institute of Technology,** ***New Jersey*** **Aug 2022 – May 2024**

*Masters of Science in Data Science* *GPA: 3.8/4.0*

Coursework: Time Series Analysis, Deep Learning, Applied Machine Learning, Statistics Methods, Optimization Methods

**University of Mumbai,** ***India*** **Aug 2018 – May 2022**

*Bachelor of Engineering in Computer Engineering* *GPA: 3.82/4.0*

Coursework: DBMS, Data Structures, Data Analytics, Software Engineering, Object-Oriented Programming

## Technical Skills

Python, R, SQL, Postgres, Docker, Spark, Airflow, Kafka, Redis, Snowflake, ETL/ELT, Tableau, Interpersonal skills

## Experience

**Minerals Technologies Inc, Bethlehem, PA** **Jun 2023 – Dec 2023**

*IT Intern -* ***MTI Oracle***

- Collaborated with the **Oracle Analytics** team of the USA, Europe and India to migrate client data from **Oracle E-Business suite** to **Oracle Analytics Cloud(OAC)**.
- Developed **composite visualizations** of financial portfolio on OAC dashboard **(Sales Data Mart and ADW Sales Voice - Datasets)** transacting CY expense of $893 Millions, PY expense of $629 Millions and 42% Change expense.
- Revamped data project with OAC's **Machine Learning module**, reducing duplicate data by 40% and facilitating timely report and dashboard creation.
- Automated incident and request handling of company's **ServiceNow Chatbot**, reducing employee manual work by 35% and improving ServiceNow accessibility by 50%.
- Mastered and deployed **Java Scripts** using in-built **Natural Language Processing(NLP)** techniques to improvise plugins(ITSM, Conversational Appointment booking).

## Projects

**Twitter Stock Forecasting - ARIFMA, GARCH** | *R, Time Series Analysis, business intelligence* **Academic Project**

- Conducted **EDA, Residual Analysis** on Twitter Stock dataset to verify if series is seasonal and stationary or not by plotting **ACF**, **PACF, QQ** graphs and took their first differencing. Obtained p-value $< 0.05$.
- Implemented ARIFMA, GARCH, GJR-GARCH models to compare **AIC, MSE, Log Likelihood** scores. **gjrGARCH(1,1)** was the best model with Loglikelihood 12370.83. Forecasted stock price for next 60 days.

**Pyspark weather API data streaming and analysis** | *Python, Machine Learning, SQL* **Academic Project**

- Engineered a **PySpark** streaming application integrating **SQL** and machine learning algorithms to dynamically fetch weather data from custom-selected locations and dates via a **weather API**(Processed over 25 fields).
- Enhanced functionality with **PyML operations** to compare weather data from two locations simultaneously, aiding event scheduling teams and skincare brands in strategic decision-making.

**Grocery Store Optimization** | *Python, A/B testing, Causal Inference, customer journey* **Academic Project**

- Conducted **Z-test**,**Chi-Square** test to analyze significant difference between sign up rate for cheaper expensive mailer of delivery club. Performed **ANOVA** to find p-values of all columns(response variable as Sales column).
- Obtained 96% loyalty score accuracy, 93% behavior classification with **Regression** algorithms, significantly transforming store layout and marketing. Applied causal inference to determine relationship between customer loyalty and delivery campaign.

**Image Search Engine** | *Python, Deep Learning* **Academic Project**

- Utilized pre-trained **VGG16** network, modifying its architecture by replacing the final **MaxPooling** layer with **Global Average Pooling** layer to generate single feature vector for each image, which facilitates image similarity.
- Extracted **feature vectors** from pre-processed base-set of 300 images are used to find similarities using **Cosine Similarity** when new search image is processed; system identifies and displays the most similar images with corresponding similarity scores.

**Product Sentiment Analysis of iPhone 15** | *Python, NLP, predictive models* **Academic Project**

- Analyzed sentiments from 20K tweets with **VADER/TextBlob**, scoring 0.09/0.12, and used **LDA** for topic modeling (perplexity 1066), boosting marketing insights. Utilized **Twitter API** to scrape tweets and created a **Semantic Network** to identify the most prominent positive & negative words associated with product.

## Publications and Courses

**Tableau Desktop Specialist(Salesforce)**

**Dr.Megharani Patil, Vatsal Kadakia, Yash Gupta, Harsh Doshi(2022); Automated Healthcare System using Gamification and AI based Chatbot; Computer Vision (Applications of Visual AI and Image Processing),vol 15,De Gruyter**

**Mendon, A., Patil, M., Gupta, Y., Kadakia, V., Doshi, H. (2023). Automated Healthcare System Using AI Based Chatbot. In: Balas, V.E., Semwal, V.B., Khandare, A. (eds) Intelligent Computing and Networking. Lecture Notes in Networks and Systems, vol 632. Springer, Singapore**