

# KHUSHAL SHARMA

(716)-547-9553 | [email](#) | [LinkedIn](#) | [github](#) | [kaggle](#) | Buffalo, NY

## EDUCATION

### **Master of Science in Engineering Science**

**Jan 2022 - May 2023**

University at Buffalo, The State University of New York | Buffalo, NY

Courses: Artificial Intelligence, Big Data Analytics, Computer Vision & Image Processing, Data Structures & Algorithms, Deep Learning, Machine Learning, Natural Language Processing, Robotic Algorithms

### **Bachelor of Science in Mechanical Engineering**

**May 2016 - May 2020**

Gujarat Technological University | Ahmedabad, India

## EXPERIENCE

### **AI/ML Software Engineer, Ilibrium | Vancouver, WA**

**Apr 2023 - Present**

- Led integration of Text-to-Speech, Speech-to-Text, Text-to-Image APIs enhancing website features, increased user engagement by 30%.
- Deployed AI solutions on AWS infrastructure including, API gateway, Lambda, S3 for improved scalability, serverless compute, storage.
- Implemented real-time ETL pipeline, leveraging Apache Kafka, enabling doctors to push/pull patient's CT scans, X-rays, reports, more.
- Employed combination of differential privacy, data masking techniques safeguarding patient's health records while ensured data privacy.

### **AI/ML Software Engineering Intern, Revolutionary Integration Group Inc. | Bethany, CT**

**Feb 2023 - Apr 2023**

- Developed biometric authentication application for periocular region of face, leveraging AWS Sagemaker, EC2, EMR, S3 services.
- Improved testing accuracy, 84% to 95% via transfer learning (VGG16, BlazeFace), A/B testing, data augmentation, hyperparameter tuning.
- Collaborated with cross-functional team of 6 to orchestrate model deployment on AWS infrastructure. Integrated flask and django for user interface. Utilised git for version control.

### **Teaching Assistant for MAE - 376, Applied Mathematics, University at Buffalo | Buffalo, NY**

**Sep 2022 - Dec 2022**

- Oversaw 300+ students, managing grading homeworks, assignments, exams, provided constructive feedback to enhance performance.
- Facilitated student learning by conducting problem-solving sessions and held office hours to address queries effectively.
- Organised study materials, collaborated with the professor to ensure the maintenance of academic integrity.
- Fostered group activities, discussions, created dynamic and engaging classroom environment for students in applied mathematics.

### **Junior Data Scientist, Evox Systems Pvt Ltd. | Mumbai, India**

**Jun 2020 - Dec 2021**

- Constructed end-to-end ETL pipelines on Azure Databricks using pyspark, scala, python, integrated PostgreSQL systems, leveraging a data lake for scalable data transformation for huge chunks of data. Orchestrated seamlessly using Azure Data Factory.
- Executed SQL queries, crafted interactive BI dashboards (PowerBI, Tableau, Excel), back data driven decision making via business reports.
- Worked with a cross functional team of 6 to perform customer churn, used logistic regression, XGBoost, random forest, SVM, Bernoulli Naive Bayes, achieved ROC-AUC score to evaluate accuracy of 93%, A/B testing to further boost accuracy up to 96%.
- Implemented CI/CD processes, deployed models for estimating customer segmentation using unsupervised techniques like k-means, mean-shift clustering, evaluated macro average F1 score of 92%. Used AWS sagemaker for deployment.

## PROJECTS

### **Airlines-Cancellation-Analysis** | Language: Python, PySpark, Scala | Analysis: MySQL | Cloud: Azure | ETL: Databricks | Tool: Google colab

- Extracted data from Kaggle's API, loaded onto Databricks-File-System (DBFS) from Azure CLI. Deployed 2 spark jobs to perform transformations on Databricks Compute, repartitioned and loaded transformed results in parquet format on DBFS.
- Utilised SMOTE technique to handle major class imbalance, build XGBoost, Bernoulli Naive Bayes, Random Forest to predict cancellation or not. Monitored ROC-AUC score, concluded Random forest with best 98% F1 score on each class.
- Collaboratively contributed to kaggle open source: [Data-preprocessing](#), [Classification-Models](#), used git as version control.

### **Stock Price Prediction & Sentiment Analysis** | Language: Python | Analysis: MySQL | Tool: Google colab | pytorch, scipy, nltk, transformers

- Evaluated LSTM-based time series models, with and without news articles to predict S&P-500 listed stock prices, leveraged sentiment analysis using Huggingface's Roberta API for computing compound scores for news articles published by WSJ, Fortune, Reuters & more.
- Implemented unique data splitting strategy, by dividing dataset to 10 days timestamp, employed first 9 days for training and final day for testing, ensuring sequential order for model training and evaluation.
- Achieved 1.5% improvement in accuracy when trained with news articles' compound scores, resulting in testing accuracy of 97%, Visualised and analysed true vs predicted stock prices to gain actionable insights.

## SKILLS

**Programming Language & Tools:** Python, C++, MATLAB, R, Java, Jupyter Notebook, Google Colab, Spyder, VS Code, RStudio

**Web Development, Frameworks & API:** HTML, CSS, Django, Flask, JavaScript, Kaggle API, React.JS, Node.JS

**Database, Version-control & OS:** SQL, PostgreSQL, SQLite3, MySQL, MongoDB, Git, LINUX, UNIX, Windows, MacOS

**Extract-Transform-Load:** Azure Databricks, Azure CLI, Azure DBFS, Azure Data Factory, AWS Elastic MapReduce, AWS S3, AWS Glue

**Big Data Analytics & Containerization:** Hadoop, Pyspark, Scala, Docker

**Visualization & Cloud technologies:** PowerBI, Tableau, Excel, Microsoft Azure, Amazon Web Services (AWS)

**Libraries:** Numpy, Pandas, Matplotlib, Seaborn, SciKit Learn, Tensorflow, Keras, PyTorch, Transformers, OpenCV, NLTK, OS, Glob.

## CERTIFICATION

- AWS Certified Cloud Practitioner (CLF-C02): [AWS Digital Badge](#)

**Oct 2023**