

# Murtaza Ali

8732882501 | Canada | murtaza.hussain2501@gmail.com | linkedin.com/in/murtaza0125 | github.com/murtaza0125 | kaggle.com/murtaza0125  
website: www.murtaza.info

## Summary

---

Highly analytical and experienced Data Scientist well-versed in modelling end-to-end data pipeline, architecture and overcoming complex optimization. Proficient in cloud apps, data processing and mining algorithms, data visualization as well as scripting languages including Python, SQL, and R. Willing to relocate and make a positive impact in an organization with my skills.

## Skills

---

- Skilled and certified in data wrangling, visualization, machine learning, natural language processing and interacting with cloud-based products with proficiency in Python, SQL, Hadoop, Pig, Spark (MLLLib), Tableau, R, Statistics
- Hands on experience with TensorFlow & Keras, Pandas, Matplotlib, Seaborn, Scikit-Learn, NLTK, BeautifulSoup and others
- Additionally skilled in No SQL, R, Tableau, Time Series Forecasting, Natural Language Processing and Recommendation Engine

## Experience

---

### Data Scientist

Sep 2022 - Current

*Securitas Technology, Ottawa, Ontario*

- Develop effective dashboards to streamline product development which improved customer satisfaction score by 40%
- Built data pipeline via MS SQL and Python for customer service case data from multiple data source, trained gradient-boosting predictive models to reduce case escalations by 30%

### Senior Product Analyst

Sep 2019 - Aug 2022

*Stanley Healthcare, Ottawa, Ontario*

- Implemented CAPA and performed Root Cause Analysis on common issues to reduce their occurrences by 95%
- Used FMEA analysis to prioritize customer stories to align business needs and improve customer satisfaction
- Revamped the technical training program and delivered it for new employees and dealers to improve the effectiveness yet reducing the onboarding time to a week from 14 days improving the productivity of the organization

### Business Data Analyst

Aug 2013 – Aug 2018

*Vedanta Resources, Silvassa, DNH, India*

- Developed and validated data visualization and reporting dashboards.
- Identified opportunities in business efficiency improvements through application of data science using Excel and Minitab
- Analyzed data and implemented various award-winning Continuous Improvement Projects which saved millions of rupees for the plant.

## Projects

---

### Used NIDDK research data to accurately predict diabetes in a patient

Sep 2022

*Toolkit: Python, GridSearchCV, AdaBoost, sklearn, seaborn*

- Optimized hyperparameters and then evaluated various ML models using standard validation techniques and evaluation parameters. The AdaBoost Classifier was identified as the champion model with 72% F1 score and 78% ROCAUC score

- Described the relevant relationships, correlations, and properties of data with various charts on a Tableau dashboard

### Reduced the time a Mercedes Benz spends on the test Bench using Machine Learning

Jan 2022

*Toolkit: Python, sklearn, PCA, XGBoost*

- Selected relevant features by variance and did dimensionality reduction using PCA
- Calculated expected time on the test bench using XGBoost Regressor to help optimize the test process

### Modeled housing loan data to predict risk of applicant loan default

Jun 2022

*Toolkit: Python, ANN, Keras and TensorFlow, SMOTE, ROC\_AUC curve*

- Performed feature engineering to select relevant features from huge number of columns and balanced it with SMOTE
- Trained and optimized an ANN to model the data and achieved 92% roc\_auc score

### Created world expansion strategy for an insurance company using Tableau

Jul 2022

*Toolkit: Excel, Tableau*

- Connected and blended diverse datasets and created calculated values needed for BI analysis
- Facilitated decision making using dashboard with interactable geographic maps, web data, KPI table, growth indicators

### Optimized productivity by Data Analysis on NYC 311 calls requests

*Toolkit: Python, datetime, matplotlib, pandas, NumPy*

- Performed a statistical test to evaluate the Null and Alternate Hypothesis with corresponding p value
- Comprehended whether average response time varies for complaint types. And if service request and location are related

### Improved overall performance by extracting information from employee database using My SQL

Jan 2022

*Toolkit: My SQL, concat, join, if-else, union, indexing*

- Facilitated performance mapping and ensured standards for HR from extracted employee ratings and salaries
- Calculated appropriate employee bonuses and training needs from the data to raise the organization performance

## Education

---

**MEng:** Engineering Management – University of Ottawa, Ottawa, ON – GPA 9.0

Sep 2019

*Relevant Coursework: Applied Data Science, Data Science for Managers with R, Market Trends*

**Data Scientist:** Master's Program – Data Science – Simplilearn in collab with IBM

Sep 2022

*Relevant Coursework: Python, R, My SQL, Machine Learning, Deep Learning, Tableau, Statistics*

**Big Data Hadoop and Spark Developer** – Simplilearn

Dec 2022

*Relevant Coursework: HDFS, YARN, Apache Hive, NoSQL, Spark, MLlib, GraphX*

**Natural Language Processing** – Simplilearn

Feb 2023

*Relevant Coursework: NLP and Speech Recognition with Deep Learning, NLTK, TextBlob, spaCy*