

DhyanKumar Shah

Data Scientist (Machine Learning, Deep Learning, NLP, Python, AWS)

📍 Toronto, Canada

🐙 github.com/dhyan1999



+1(807)358-5360



[LinkedIn](#)



dhyankumar05051999@gmail.com



SUMMARY

- Data Scientist with strong math background and **2+ years** of experience using **predictive modeling, data processing, and data mining** algorithms to solve challenging business problems. Involved in Python open-source community and passionate about **deep reinforcement learning**.
- Strong foundation of **statistical modelling** applied **machine learning**, recommendation systems, **Extract, transform, and load (ETL)** and **Data Visualizations** which aid me in tackling various challenges that come with data processing.
- I am extremely passionate about working with huge chunks of data and harnessing meaning from it. Working under pressure, **solving challenges**, making **business driven decisions** are some things I excel at.
- Improved accuracy of credit default prediction model from **67.4%** to **82%**, which helped our client increase their revenue by **~18%** Year on Year.
- Experienced in real time data **reporting** and **presenting** insights using **dashboards** and stories via Advanced Excel, **Tableau**, Microsoft **PowerBI**.
- Demonstrated skills in data handling using **Microsoft SQL Server, MySQL, Oracle, MongoDB**.
- Experienced in building **sentiment analysis** models using **Python** and **Spark**. Skilled at **scraping** internet sources and developing customer support internet bots.
- Proficient in utilizing cloud platforms such as AWS (**Amazon Web Services**) and Azure (**Microsoft Azure**) to store, process, and analyze **large-scale datasets**.



TECHNICAL SKILLS

Skill	Technology worked on
Programming /Scripting	Python (PyData Stack), PySpark, R, C, C++, JavaScript
Data Visualization	Tableau, Power BI, Advance Excel, Matplotlib, Seaborn
Machine Learning	Feature Engineering, Exploratory Data Analysis (EDA), Statistical modelling, Natural language Processing, NumPy, Pandas, Scikit-learn, NLTK, Spacy, Gensim, Keras, TensorFlow, Image Classification, Object Detection
Deep Learning	ANN, CNN, RNN, LSTM, BERT, Transfer Learning
Version Control	Git, Microsoft TFS
Big Data & Cloud	Databricks, Apache Spark, Hadoop Ecosystem, AWS (Amazon Web Services), Microsoft Azure
Database	MySQL, MSSQL, MongoDB, PostgreSQL
Operating system	Windows, MacOS, Linux



EDUCATION

- **Master of Science (Computer Science) 2021 – 2022**
Lakehead University, Thunder Bay- Ontario (Canada)
Graduated Master's degree with Distinction 3.62/4.00 GPA
- **Bachelor of Technology (Computer Engineering) 2016 - 2020**
Indus University, Ahmedabad- Gujarat (India)
Graduated Bachelor's degree with Distinction 3.72/4.00 GPA



PROFESSIONAL EXPERIENCE



Genious Infotech *(September 2022 to Present)*

- **Role and responsibilities: Junior Data Scientist**
 - Responsible for improving model accuracy of existing credit default prediction model.
 - Develop action plans to mitigate risks in decision making while increasing profitability by leveraging data science.
 - Drive the interaction and partnership between the managers and ensure active cooperation in identifying as well as defining analytical needs and generating the pull-through of insights with the business.
 - Prepare the predictive models for deployment in a AWS production environment. Work with software engineers to integrate the models into operational systems and establish monitoring mechanisms to track their performance and recalibrate them if necessary.
 - Constructed a language model for multi-class classification of Hate Speech (Religion, Racism, Cyberbullying, Feminism, General Hate) using Bert T5 model and achieved F1-score of average 75 on 0.15 million tweets.
 - Applied Twitter API and Beautiful Soup to scrape 30% of the data and manually annotate it.
 - Implemented Word, Phrase, and Document Level Data Augmentation and increased hate data by 45%.
 - Leverage the use of Topic Modeling, BOW, Glove, Word2Vec techniques for feature extraction and classification.
- **Tools Used:** Python, PySpark, XGBoost, AWS, BertT5, Text Augmentation, Classification, Model Cleaning and Preprocessing



GVT Software Solution *(January 2020 to December 2020)*

- **Role and responsibilities: Data Analyst**
 - Collaborated with the data engineering team to extract, preprocess, and optimize data from various sources, leveraging AWS services such as Amazon S3, as well as Hadoop's distributed processing capabilities.
 - Conducted exploratory data analysis (EDA) to identify trends, patterns, and outliers, providing valuable insights to inform strategic decisions.
 - Designed and created interactive dashboards using Tableau, showcasing key performance indicators (KPIs) derived from AWS-hosted data sets processed through Hadoop, enhancing data-driven decision-making.
 - Assisted in the development of machine learning models to predict customer behavior, resulting in a 23% increase in accuracy compared to the existing model.
 - Participated in cross-functional team meetings to communicate findings and recommendations, contributing to the company's data-driven culture.
- **Tools Used:** Python, SQL, GIT, Tableau, Apache PySpark and Kafka, AWS, Hadoop, Advance Excel, Scikit Learn



Carleton University *(June 2019 to July 2019)*

- **Role and responsibilities: Data Analyst (Student Exchange Program)**
 - Executed projects related to Data Analysis with Matplotlib, seaborn, regression, classification and clustering. My final assignment involved analyzing a data collection of all FIFA-19 players, and I achieved a 95% accuracy rate.