

EDUCATION

Indiana University, Bloomington

December 2022

Master of Science, Data Science

GPA: 3.7/4

Coursework: Statistics, Machine Learning, Deep Learning, Big Data, Explainable Artificial Intelligence, Advance Database Concepts

University of Mumbai, India

May 2019

Bachelor of Engineering, Computer Science

GPA: 3.8/4

TECHNICAL SKILLS

Programming Languages: Python, R, Java, Pyspark, HTML-CSS, PHP

Databases and Big Data: SQL, MySQL, NoSQL, Hadoop, Spark, Microsoft Azure, AWS, Scala, GCP

Statistical Analysis: Hypothesis Testing, ANOVA, Chi-Square, Predictive/Descriptive Analytics, A/B Testing

ML Algorithms: Regression, Classification, Clustering, Decision Trees, Random Forest, XGboost, LightGBM, KNN

DL Algorithms: CNN, RNN, LSTM, GRU, MLPs, GANs, Artificial Neural Network, Siamese Network

Analytical Tools: Microsoft Excel, Tableau, PowerBI, VBA

Tools & Libraries: NumPy, Pandas, Matplotlib, Selenium, OpenCV, Scikit-learn, PyTorch, TensorFlow, Keras, Docker, Git

WORK EXPERIENCE

Research Assistant

June 2021 – Present

Indiana University, Bloomington

- Collected data from 4 different data sources and analyzed the student participation in course discussion threads along with academic participation of around 200 students.
- Reduced the manual analysis time of course participation by 80% through the statistical analysis of huge course data.
- Discovered insights by applying statistical analysis, dimension reduction, variable re-identification, and synthetic data generation. Used NLP techniques to study the semantics and perform topic modelling on huge university data consisting of 100,000 data points.
- Built an Artificial Neural Network using Tensor Flow to identify the student's probability of retention based on their library use.

Data Engineer

December 2019 – December 2020

Course5 Intelligence, Mumbai

- Robotic Process Automation for an e-commerce website, setting prices and promotions, and other manual time-consuming processes/enablers on both websites, desktop apps, and different portals.
- Undertook an entire process for automated product inventory update and health checks for the website using Java and Selenium, providing end-to-end solutions and possible ways to optimize the process.
- Formulated real-time solutions for migrating production database from an archaic wireframe to Google Big Query.
- Prepared interactive visualizations on 20+ dashboards in Google Data Studio for seamless engagement of clients.

Developer

August 2019 – November 2019

Mindchamp, Mumbai

- Hosted organization and course data on website using Stremlit and AWS EC2 Instance.
- Improved content delivery and decreased end-user latency by 60% of web application using Amazon CloudFront.
- Created organized lesson plans to teach simplified, fun computer coding to school kids with the help of block-based coding platforms such as Code.org and MIT Scratch.

ACADEMIC PROJECTS

Clothing Similarity Understanding with Deep Learning

October 2021 - December 2021

(InceptionV3, VGG16, ResNet50, KNN and Siamese Network)

- Find the most suitable clothing matches for the current selection and choice of clothing using existing pre-trained Deep Learning based feature extractors. Used Siamese Network to find the similarities between the images across various categories.
- The Siamese Network incorporated with the feature maps extracted from VGG16 was able to fully capture the visual similarity and clothing category matching the real intentions of the user.

E-Commerce Recommendation System

March 2021 - May 2021

(KNN, Stochastic Gradient Descent, ALS and Neural Collaborative Filtering)

- Presented the workings of four different models under the collaborative filtering approach of recommendation systems.
- Demonstrated how to evaluate the model's performance using different metrics such as hit ratio (91%) and mean absolute error (0.39).

Chicago Crimes Data Analysis

August 2020 - October 2020

(Logistic Regression, Exploratory Data Analysis, Predictive Analysis)

- For the year 2019, based on our analysis only 19 % of the crimes that were reported, were arrested. Out of all the crime types 'theft' was the highest committed crime with 20.85 % of the all the crimes reported, out of which only 5% were marked arrested.
- There was a significant amount of interaction between the type of crime and location with respect to the target variable arrest made. Different GLM models were used, and the best model was chosen based on the comparison with the original data trend.