HANY RAZA

hanyraza111@gmail.com | Boston, MA |https://www.linkedin.com/in/hany-raza-lkdin| 6178240077

Available: Dec 2021 – Dec 2022

EDUCATION

Northeastern University **Boston, MA** **Jan 2021- Dec 2022**
Khoury College of Computer Sciences **GPA: 3.78/4**
Master of Science in Computer Science
Relevant Coursework: Machine Learning, NLP, Foundations of Artificial Intelligence, Algorithms

Sardar Patel Institute of Technology **Mumbai, India** **Sept 2020**
Bachelor of Engineering in Information Technology **GPA: 8.33/10**
Relevant Coursework: Data Analytics, Data Warehousing & Mining, Soft Computing, Mathematics & Statistics, Advanced Database Systems, Object Oriented Programming, Enterprise Resource Planning, User Experience Design, IOT, Software Engineering, etc

TECHNICAL KNOWLEDGE

- **Languages & OS**: Python, Java, R, C/C++, SQL, Git/GitHub; Windows and Linux
- **Frameworks & Libs:** NumPy, Pandas, Matplotlib, Keras, TensorFlow, OpenCV, SciPy, Sklearn, Seaborn, Hadoop, Django
- **Databases**: MySQL, MongoDB, SQLite**| Cloud**: GCP, Heroku**| Dev Tools:** Orange, MS Excel, Tableau, Anaconda
- **Certifications and Courses:** Machine Learning, Deep Learning, Maths for Machine Learning Specialization, Applied Algorithms & Data Structures (Morgan Stanley), Data Analytics in R, Data Visualization with MS-Excel & Tableau, Google Cloud Platform
- **Soft Skills:** Effective Communication & Presentation, Strategic Thinking, Data-driven Decision Making, Researching, Planning

PROFESSIONAL EXPERIENCE

Khoury College of Computer Science **Boston, MA** **Feb 2021 – Apr 2021**
Graduate Teaching Assistant - *CS 2810: Mathematics of Data Models under Prof. Kevin Gold*
- Course focused on mathematic principles for Data Science/ML like linear algebra, multivariable calculus, statistics, etc

Vamrr Technologies Pvt. Ltd. **Mumbai, India** **Aug 2020 – Oct 2020**
Technology Ninja - *Product: Appointment Booking Intelligent Chatbot Software*
- Used NLP and Deep Learning for training chatbot's responses as per user's inputs and tuned hyperparameters in RASA
- Used MySQL database in back-end and wrote CRUD queries for customer data and integrated it with the software
- Also, Integrated Google Calendar and Textlocal service, finally increasing customer retention and appointment bookings by 15%

Prosares Solutions Pvt. Ltd. **Mumbai, India** **Dec 2019 – Jan 2020**
Machine Learning Intern - *Project: Automatic and Specific Image Tagging using Machine Learning Techniques*
- Researched various technologies to classify images for proper storage and expeditious retrieval and performed pre-processing operations over it for image denoising, enhancement, transformation, etc.
- Implemented Convolutional Neural Networks (CNN) to accurately detect specific features of different objects in each image using Keras and TensorFlow libraries from Python, automating and enabling a much faster and efficient way of tagging images.

Sardar Patel Institute of Technology **Mumbai, India** **July 2019 – May 2020**
Undergraduate Teaching Assistant – *Python Programming and Object-Oriented Programming (Java)*

ACADEMIC PROJECTS https://github.com/hany110

Predicting and Analyzing Churned Users
- Used Light GBM model and calculated accuracy to be 94.63% with AUC score 0.8, further understanding "Why users churned?"
- Extracted relevant features based on defined problem and wrangled, cleaned and processed it to obtain target feature variable (churned users) for training dataset. Sampled another feature-set and used SQL queries to fit target variable over it
- Performed data pre-processing, analysis and filtered top 25% features by random forest due to low correlation between features

Cancer Prediction
- Prediction results pronounced XGBoost as best, having the maximum accuracy of 95.32%, followed by SVM and Random Forests.
- Compared Logistic Regression, Naïve Bayes, SVM, XGBoost, Random Forests and KNN to find best performance using F-1 score.
- Tuned hyperparameters using GridSearchCV and cleaned, wrangled, explored data using Pandas, Numpy and Matplotlib

Predict Future Service Demand
- Analyzed time-series possibility, processed data and created new features. Explored data by countplot & distplot from seaborn.
- Assessed correlation using heat map, selected top 25 features using SelectKbest from sklearn and compared random forest, SVM, KNN and XGBoost using RMSE and MAE. XGBoost gave best accuracy of 93.87%, giving good prediction for future demand

Weather Forecasting
- Employed web scraping to gather raw weather data and Implementing time-series analysis to forecast the weather ahead
- Analyzing feature dependency, smoothing, decomposing in trend, seasonality, noise to effectively visualize and understand data
- Comparing SARIMA, ARIMA, LSTM models and using Numpy, Pandas, Matplotlib, Beautiful Soup, Seaborn, Statsmodels, etc.

Product Review Sentiment Analysis
- Performed cleaning and analysis over raw product data and used Logistic Regression acquiring an accuracy result of 93.8%
- Employed feature engineering using correlation analysis, backward selection with p-value and visualized results using Matplotlib

Finding Frequent Itemsets
- Analyzed, wrangled and explored 300,000 global retail sale data of company, and employed Apriori Algorithm over it, tuned its parameters and visualized the frequent item clusters using scatter plot, which aids in predicting customer's next product choice