ROHIT RAJESH KHATU.

Boston, MA | (617) 818-3265 | github.com/rkhatu97 | khatu.r@northeastern.edu | linkedin.com/in/rkhatu97

EDUCATION

Northeastern University — Boston, MA
Master's in Data Analytics, concentration Machine Intelligence GPA: 3.93 — *December 2021*
Relevant Courses: Probability and Statistics, Big Data, Data Visualization, Intermediate Analytics, Data Mining.

Mumbai University, Saraswati College of Engineering — Maharashtra, India
Bachelor of Engineering, Computer Engineering — *May 2018*
Relevant Courses: Operating Systems, Big Data, Soft Computing, and Data Structures.

TECHNICAL SKILLS

Programming Languages: Java, Python, R, SQL, JavaScript, Shell Scripting.
Web Development: HTML, CSS, Bootstrap. Apache Tomcat.
Database/ Big Data Technologies: MySQL, MariaDB, Hadoop, HDFS, Sqoop, Hive, MapReduce, Spark.
Cloud Technology and Automation: Docker, AWS (EC2, S3), Azure, GCP, Jenkins, bamboo, Jira, chef.
Tools and Packages: GIT, Microsoft Office, Tableau, PowerBI, Anaconda, Scikit-learn, TensorFlow, PyTorch, NodeJS.
Statistics: Hypothesis Testing, Descriptive Statistics, A/B Testing, ANOVA, Chi-Square Test, Pearson Test.
Machine Learning Algorithms: Linear Regression, Ridge and Lasso Regression, Logistic Regression, KNN, Naïve Bayes, K-Means clustering, Apriori, Decision Tree, Random Forest, Boosting Algorithms, PCA, SVM, CNN, LSTM.

WORK EXPERIENCE

Dtonomy Inc. — Cambridge, MA
Machine Learning Software Engineer Intern — *January 2021 – June 2021*

- Refined random forest model used for pattern discovery by performing feature addition and hyperparameter tuning. Deployed the model as a service on Docker container for identifying security detection and provide analysis.
- Constructed a node-red workflow using NodeJS, JavaScript, HTML, and python to set inbox rules to block spam mails with keyword and to block sender with id, leveraging Microsoft Graph API and Azure app registration feature.
- Developed selenium script automating request and receipt of consent for scopes of graph API used to access and set the inbox rules. Created inject node in node-red to trigger workflow based on time, time zone and various data types.
- Created a PowerShell node in nod-red to run the PowerShell script. Achieved OS independence for the node by installing PowerShell dependencies into the docker container and deploying the node on top of the container.

Mindcraft Software PVT.LTD. — Mumbai, India
Associate Consultant (DevOps Developer) — *August 2018 - December 2019*

- Leveraged DevOps technology; Jenkins, docker, and chef for automated deployment of MySQL server and client on multiple servers simultaneously which reduced the deployment time to 2 Hours from 1 week.
- Automated Deployment of Loan disbursement Web application on Tomcat server for DEV/UAT/PROD environment using single Jenkins job, achieved timely deployment and testing.
- Developed Shell scripts to perform database/WAR file backup, user/database/table creation, and set user access rights as per project security guidelines. Triggered these shell scripts using cron jobs and Jenkins.
- Engineered software with help of ruby scripting for Universal Overseas Bank, Singapore. Documented POC's for the entire project, worked under Agile practices. Working Knowledge of finance and insurance sector.

ACADEMIC PROJECTS

Credit Card Fraud Detection: — November 2020

- Imported and Analyzed 700K transactions from HDFS in python using PySpark. Converted JSON to a data frame. Performed data cleaning using NumPy, pandas; EDA using matplotlib, seaborn. Visualized correlation among target and features using Corr plot. Oversampled data to handle imbalance. Standardized data using standard-scalar.
- Build GBM, Random Forest, and sequential model, sequential performed better in predicting if a transaction is fraud.

Product Review Sentiment analysis: — September 2020

- Analyzed 20 years of amazon product review using python, segregated reviews based on ratings, below 2 stars as negative, equal to 3 stars as neutral, and above 3 as positive, built bar plot to visualize data count of all three sentiments.
- Converted the data using TFID vectorizer, constructed ML models Naïve Bayes, logistic regression, and SVM, choose logistic based on metrics, saved as PKL, build web UI using flask to accurately predict the sentiment of a review.

Facial Recognition: — August 2020

- Developed a python program using OpenCV to generate train and test datasets of facial images.
- Rendered a neural network model by using the VGG16 transfer learning technique, Keras, and TensorFlow for feature extraction from images. Performed hyperparameter tunning and saved model in h5 format.
- Imported the model, provided external IP webcam support to use external devices for effective face recognition.

Crimes in Boston Data Analysis: — February 2020

- Analyzed a million entries on crimes occurred in Boston from year 2016-19 in R. Developed dashboard using RShiny to display changes in crime rate. Used ARIMA forecasting model to predict the change in crime rate for six months.

CERTIFICATION:

Neural Network and Deep Learning (*Coursera*) — March 2021