

ROHIT RAJESH KHATU.

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

EDUCATION

Northeastern University

Master's in Data Analytics, concentration Applied Machine Intelligence GPA: 3.93

Relevant Courses: Probability and Statistics, Big Data, Data Visualization, Intermediate Analytics, Data Mining.

Boston, MA

December 2021

Mumbai University, Saraswati College of Engineering

Bachelor of Engineering, Computer Engineering

Relevant Courses: Operating Systems, Big Data, Soft Computing and Data Structures.

Maharashtra, India

May 2018

TECHNICAL SKILLS

Programming Languages: SQL, Java, Python(NumPy, Pandas, SciPy), R(dplyr, ggplot2), JavaScript, Shell Scripting.

Web Development and Web Server: HTML, CSS, Bootstrap. Apache Tomcat.

Database: MySQL, MariaDB.

Cloud Technology and Automation: Docker, AWS (EC2, S3), Azure, GCP, Jenkins, bamboo, Jira, chef.

Tools and Packages: GIT, Linux, Microsoft Office, Tableau, PowerBI, Anaconda, TensorFlow, PyTorch, NodeJS.

Big Data Technologies : Hadoop, HDFS, Sqoop, Hive, MapReduce, Spark.

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.

Machine Learning Software Engineer Intern

Cambridge, MA

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 key word and to block sender with id, leveraging Microsoft Graph API and Azure app registration feature.
- Developed selenium script to automate the process for requesting and receiving consent for scopes of the 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 PowerShell script. Achieved OS independency for the node by installing PowerShell dependencies into docker container and deploying the node on top of container.

Mindcraft Software PVT.LTD.

Associate Consultant (DevOps Developer)

Mumbai, India

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. Deployment of Loan disbursement Web application on Tomcat server was automated 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 guideline. 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 entire project, worked under Agile practices. Working Knowledge of finance and insurance sector.

ACADEMIC PROJECTS

Credit Card Fraud Detection: Python

November 2020

- Analyzed transaction after importing data in python from HDFS using PySpark and performed EDA using matplotlib, seaborn, data cleaning using NumPy, pandas and standardized data using standard scalar
- Developed models using logistic regression, GBM and Random Forest to predict if a transaction is fraud.

Product Review Sentiment analysis: Python, HTML, CSS (Flask Deployment)

September 2020

- Analyzed 20 years of amazon product review, segregated reviews based on ratings that is 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 sentiment.
- Constructed ML model by first transforming the data using TFIDF vectorizer and applying Naïve Bayes, logistic regression, and Linear SVM in python, choose logistic regression as best of all based on metrics, saved model as PKL file and used to accurately predict sentiment of amazon reviews to see if a product has successfully captured the market.

Facial Recognition: Python, TensorFlow, Keras

August 2020

- Developed a python program using OpenCV to generate train and test dataset of facial images.
- Rendered a neural network model by using VGG16 transfer learning technique, keras and TensorFlow. Performed hyperparameter tuning 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: R, RShiny

February 2020

- Analyzed a million entries dataset on crimes occurred in Boston from year 2016-19. Developed heat map to show change in crime rate as per county. Used ARIMA forecasting model to predict the change in crime rate for six months.