Rohit Bhanushali

480-779-0782 | rohit.bhanushali27@gmail.com | Seattle, WA | **LinkedIn** | **GitHub** | **Website**

Seeking a dynamic role that allows me to leverage my expertise in Machine Learning and Software Development to drive growth & innovation and utilize my strong analytical & problem-solving abilities to solve complex problems & deliver solutions to contribute to the success of the organization.

EDUCATION

Master's in Computer Science - Arizona State University, Tempe, AZ, USA   **GPA: 3.8/4.0**
Bachelor's in Computer Science and Engineering - University of Mumbai, Mumbai, MH, India   **GPA: 3.7/4.0**

TECHNICAL SKILLS

Languages: Python, Java, C++, C
Databases: MySQL, MongoDB, MemSQL, PostgreSQL, Amazon RedShift
AWS: SageMaker, Step Functions, AWS Lamba, S3, EC2, CloudWatch, Athena, Glue, IAM, Data Pipeline
Azure: Azure Functions, Azure Machine Learning, Kubernetes Service, Blob Storage, Azure Monitor, Azure Data Factory
Tools: Docker, Container, Git, Azure DevOps, Jira, Azure Boards, MySQL Workbench, Excel, Tableau
Web: HTML, CSS, JavaScript, Flask, Django, RESTful APIs, AngularJS, ReactJS, D3.js, Elasticsearch, Ajax
Libraries: NumPy, Pandas, SciPy, PyTorch, Transformers, Scikit-Learn, TensorFlow, Keras, NLTK, Matplotlib, Seaborn
Machine Learning: Classification, Regression, Statistical and Predictive Analysis, MLOps, PCA, K-means, KNN, Deep Learning
Big Data: Hadoop, MapReduce, Spark, HBase

PROFESSIONAL EXPERIENCE

Software Development Engineer - Machine Learning, Amazon   **January 2021 - Present**

- Developed a deep learning-based computer vision framework using **Python, OpenCV, Spark, and ML techniques** to identify the optimal real estate placement locations in real-time by analyzing satellite and street images for traffic patterns, competition, parking spots, accessibility, and trends, resulting in a **30% increase in the success rate of real estate site acquisition**.
- Implemented large-scale data and economic modeling systems to analyze and model vast amounts of data to determine the most prime locations for new Amazon retail stores using python and machine learning algorithms like **SVM, Linear Regression, Random Forest**, and many more, leading to an **increase in in-store revenue by over 20%** within the first year.
- Optimized routing and supply chain processes using network modeling tools for a **15% reduction in transportation costs** and streamlined business requirements, **ETL pipelines**, and prototyped models to production on AWS and SageMaker, **improving workflow efficiency by 40%**.
- Developed regression-based **revenue prediction** models for new Amazon retail store types and collaborated with economists to improve model performance for the existing stores, resulting in a **25% increase in sales prediction accuracy**.
- Implemented a **PyTorch-based Anomaly Detection framework** to detect outlier real estate sites, resulting in a **20% reduction in errors** when selecting real estate sites.
- Built a framework utilizing **demand models and causal inference** to help in making better business decisions on real estate sites, pricing, and lease duration, resulting in a **20% increase in return on investment**.

Software Development Engineer - Data Science, PricewaterhouseCoopers LLP   **August 2019 - January 2021**

- Developed and implemented a novel active learning framework using **Python, PyTorch, Azure Kubernetes, and MySQL** with feedback mechanisms to significantly enhance the performance of deep learning-based anomaly detection models by iteratively updating the models with additional labeled data, resulting in a **15% increase in AUC**.
- **Boosted training speed by over 50%** by migrating to Azure and utilizing specialized compute clusters with GPU support for constructing new ML pipelines and faster training of models.
- Deployed multiple **GPU-powered models as web services on Azure Kubernetes** Service clusters using Azure ML and custom Docker images for real-time inference, **improving response time by 72%**.
- Created a time-series anomaly detection model using **LSTM Autoencoder** with PyTorch to detect anomalies in multi-dimensional ledger datasets, **reducing false positives by 20% compared to traditional methods**.
- Developed and implemented a **PyTorch**-based **Autoencoder** and DaGMM (Deep Autoencoding Gaussian Mixture Model) anomaly detection framework to detect accounting anomalies in multi-dimensional ledger datasets, **increasing the detection rate by 20%**.
- Built a python tool to optimize threshold values using the **F1 score** for capping the anomalous entries to cluster for review, **improving recall by 30%**.
- Designed, developed, and deployed well-documented, scalable, and reusable **REST APIs and Azure Functions** on Azure, allowing more efficient collaboration and easier maintenance of the models and pipelines, and **increasing developer productivity by 25%**.

Software Developer - Machine Learning Intern, CYR3CON   **May 2018 - August 2019**
- Led the design and implementation of ETL pipelines and ML models on Amazon Web Services (AWS) utilizing **Apache Airflow, Python, Selenium, MongoDB, and CI/CD practices** to automate unit and integration testing and training of new ML models, resulting in a **25% improvement in efficiency**.
- Created an advanced ML model to predict cyberattacks by analyzing and classifying the web data to vulnerabilities using techniques such as **web mining, topic modeling (Doc2Vec and LDA)**, and **classification algorithms like Random Forest and SVM**, leading to a **21% increase in the accuracy of potential cyberattack prediction**.
- Built interactive dashboard using **REST APIs, AngularJS, Flask, and D3.js** to display results and analysis from ML models, resulting in a **30% increase in user engagement**.
- Developed a python-based tool to detect potential malware using web mining and Word2Vec, **improving the detection rate of potential malware by 15%**.
- Analyzed CVEs to identify key malware-driving features using regression and ensemble techniques, **increasing the accuracy of identifying key features by 27%**.

Software Developer – Research Assistant, Data Mining and Machine Learning Lab   **February 2018 - December 2018**
- Successfully published a **research paper** centered around deep learning techniques to detect anomalies in attributed networks titled '**Deep Anomaly Detection on Attributed Networks**' at SDM19 [**Paper**], **receiving 210+ citations**.
- Developed a python-based software solution to **assist in disaster management and response** using machine learning algorithms like **Causal Forest, Random Forest, Linear Regression**, and social media data to predict user mobility and causal relationships between users during disasters.
- Built a deep learning model for measuring the **relative engagement of YouTube videos** and predicting trend duration using **Python, NoSQL, TensorFlow, NTLK, and YouTube API**.
- Implemented a system to **maintain the integrity of information** and prevent the spread of fake news by detecting and analyzing bot features in crawled tweets using **Python, MongoDB, AdaBoost, and Decision Tree Classifier**.
- Created a **real-time Twitter insights tool** that tracks, analyzes, and understands activity on Twitter by crawling millions of tweets using **Python, Flask, Twitter API, MongoDB, and REST APIs**.
- Developed a python-based interactive tool to visualize topics and keywords and explore the content of tweets using **D3.js, Plotly, LDA, and Doc2Vec**.

PROJECTS

GeoSpatial Computation Using Spark
- Developed an advanced geospatial analytics solution using **PySpark, Hadoop and Getis-Ord score on GCP** to analyze the New York Yellow Taxi dataset.
- Identified high-demand areas for taxi pickups through the analysis of hot zones, as well as the computation of Getis-Ord scores.
- Measured the effectiveness of the solution using metrics such as the accuracy of hot zone and hot cell identification, response time for the analysis and cost savings resulting from the optimized operations on **GCP**.

LeadLine: Interactive Tool for Visual Analysis of Textual Data
- Created an advanced analytics system, utilizing **Flask, Python, MongoDB, D3.js and LDA** to enable interactive visualizations of early event detection on news and social media data for easy real-time exploration and insights.
- Metrics such as event detection accuracy, response time for event detection, click-through rate, and user engagement with the visualization were used to measure the effectiveness of the system.

Sensor Fusion Calorie Detection and Diet Assessment
- Developed an innovative Android application that utilizes advanced computer vision techniques including **food segmentation, Mask R-CNN, and SVM** to accurately detect calorie intake using both thermal and color camera sensors.
- Application utilizes contour detection and hierarchical image segmentation in **Matlab** to perform **border detection, and grey/thermal image projection on edgeROF** to increase accuracy and food detection capabilities.
- System was built using a combination of **Android, Matlab, SVM, Mask R-CNN, Python, and Java and deployed on AWS**.

Analysis towards Domain-Independent Anomaly Detection
- Developed an anomaly detection system using machine learning algorithms including **k-NN, SVM, Multivariate Gaussian Distribution, and Neural networks** to prevent unauthorized transactions using **Python, TensorFlow, HTML5, and CSS**.
- Utilized various performance metrics such as **precision, recall, and F1-score** to evaluate the model performance and select the best algorithm for anomaly detection.

Advanced Recommendation System
- Developed a comprehensive **web and mobile-based recommendation system** that integrated collaborative filtering, clustering techniques, and sentimental analysis to provide a personalized product, music, and movie recommendations.
- System was built using a combination of **Java, Python, Android, J2EE, HTML5, CSS, SQLyog, and JavaScript**.
- Metrics such as click-through rate, conversion rate, and user engagement were used to evaluate the system's performance.