Rohit Bapat

(812) 606-7949 rohbapat@gmail.com rohitbapat.github.io bit.ly/tableauresume linkedin.com/in/rohitcbapat

EDUCATION

Indiana University, *Master of Science in Data Science*, Bloomington, IN, United States — *Aug 2018 - May 2020*
Coursework: Machine Learning, Statistics, Search, Social Media Mining, Big Data, Data Visualization, Deep Learning — *GPA: 3.77/4.0*

University of Pune, *Bachelor of Engineering in Computer Engineering,* Pune, India — *Aug 2013 - Jun 2017*
Coursework: Data Mining Techniques, Business Analytics and Intelligence, Data Structures and Algorithms — *GPA: 3.60/4.0*

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, Javascript
- **Database:** MySQL, MongoDB
- **Applications:** PowerBI, Tableau, Apache Storm, Apache Spark, GIT, Jenkins, Elasticsearch, Gephi
- **Frameworks:** Pandas, NumPy, sklearn, matplotlib, seaborn, tensorflow, bokeh, Python Flask, Django
- **Machine Learning:** KNN, Adaboost, Random Forests, SVM, K–Means, Logistic Regression, Light GBM, XGBoost, Statistics, Deep Learning

PROFESSIONAL EXPERIENCE

Kenzen (kenzen.com – Heat stress safety device), ***Research Data Scientist*** — *Jul 2020 - Present*
- Designed the database schema to migrate the existing FTP based data to structured tables in MySQL for improved scalability.
- Created PowerBI dashboard with interactive visuals and drill through for improvement in data visibility over static PDF reports.
- Decreased the average time for client reporting from 1 week to 3 hours with PowerBI dataflow for the dashboard.

Myxx (myxxrecipes.com – Recipe based one-click shopping service), ***Data Science Intern*** — *May 2019 - May 2020*
- Boosted the ingredient to store product conversion rate from 55% to 85% with scraper and regex improvements in Ruby.
- Implemented a new cosine similarity based Elasticsearch scoring method for consistent ingredient mapping in MongoDB.
- Introduced a multilabel classification approach to tag recipes with diet and ingredient keywords to avoid manual efforts and errors.

TIBCO Software, *Junior Consultant* — *Jul 2017 - Jul 2018*
- Designed and modeled the business processes for client in telecom domain and interfaced TIBCO products with external third party resources.
- Identified escalation time points and total process execution time with role specific reports using TIBCO Spotfire.
- Successfully deployed a TIBCO BPM Instance with data and organization models on AWS EC2 to offer a sandbox mode for the client.

Persistent Systems, *Project Intern* — *Jun 2016 - Jun 2017*
- Developed a business intelligence based solution to provide reporting and insights about sales, inventory, coupon promotions and customer footfalls.
- Built a reporting system in Jasper Reports and Jasper Server, with CSV data transformation into database using Talend ETL to achieve scalability.

PROJECTS

Bloomington Hack: Let's Get Visual - Civic Data Challenge *Python | Bokeh | Tableau | Altair* — *Feb 2020*
- Explored and visualized Bloomington city data about potholes and storm water sewers to improve public complaint resolution.
- Showcased the dashboard with revised revenue distributions to audience of 50, including residents and city council members.

Sentiment based Image Captioning *Python | tensorflow 2.x |* Numpy *| MS-COCO Dataset* — *Dec 2019*
- Worked on Sentiment based Image Captioning project using the pre-trained ImageNet CNN and LSTM network.
- Compared Inceptionv3 and MobileNetv2 performances using metrics like METEOR, CIDEr, SPICE, BLEU after adding sentiment encoding.

IndyCar - Performance Analysis of Anomaly Detection Application *Apache Storm | Tableau | Python* — *May 2019*
- Successfully deployed a Storm topology using yaml files, Apache MQTT pub-sub broker, Zookeeper to analyze IndyCar race data.
- Used Apache Storm for streaming analysis and Tableau for visualizations and insights.

Image Orientation Classification *Python | Pandas | Numpy | matplotlib* — *Dec 2018*
- Implemented Adaboost technique from scratch to identify orientation of 40,000 Flickr images.
- Achieved overall accuracy of 69.48% for test images with 4000 weak classifiers.

DonorsChoose.org Application Screening *Pandas | Python | sklearn | matplotlib* — *Nov 2018*
- Predicted application decision of DonorsChoose.org application dataset via Kaggle competition.
- Applied Natural Language Processing (NLTK) coupled with inflect libraries and textBlob packages.
- Used scikit-learn models of Light GBM (AUC 0.766) and K-Means for prediction and essay review.