Rohit Chandrashekhar Bapat

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

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 and Databases:** Python, JAVA, R, Javascript, C++, SQL, MySQL, MongoDB, Apache Cassandra
- **Applications:** Apache Storm, Tableau, Apache Spark, GIT, Jenkins, Elasticsearch, PowerBI, Google Colab, 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, Time Series.
- **Deep Learning:** (Deep, Convolutional, Recurrent) Neural Networks, Siamese Networks, Variational Autoencoders, Attention Models.

WORK EXPERIENCE

Kenzen Inc (Heat stress safety device startup), ***Research Data Scientist***, *New York City, New York* *Jul 2020 - Present*
- Developing algorithms and machine learning models for heat stress safety device.
- Created a data pipeline and database design to facilitate analytics team for reporting.
- Developed a Power BI Dashboard with data flows and MySQL Database backend to decrease reporting efforts by 50%.

Myxx Inc (Recipe based one-click shopping services), ***Data Science Intern***, *Cary, North Carolina* *May 2019 - May 2020*
- Developed Ruby and Python scripts for ingredient mapping and retailer integration.
- Boosted the website recipe conversion success rate from 55% to 85%.
- Implemented a new cosine similarity based Elasticsearch scoring technique for ingredient mapping.
- Achieved an F1 score of 0.88 for prediction of 11 tags for 6000 recipes using multilabel classification.

TIBCO Software Inc, *Junior Consultant*, *Pune, India* *Jul 2017 - Jul 2018*
- Designed and modeled the business processes for client in telecom domain and interfaced TIBCO products with external third party resources.
- Used Spotfire- TIBCO Analytic Tool to determine time required for process execution and staged reports.
- Successfully deployed a TIBCO BPM Instance on AWS EC2 with Jenkins scheduled jobs for client demos.

Persistent Systems Pvt Ltd, *Project Intern*, *Pune, India* *Jun 2016-Jun 2017*
- Developed a business intelligence based solution to provide the retail store chain with analytics based on customer footfalls, inventory management, and sales trends by generating role specific PDF reports.
- Implemented ETL phases using Talend ETL, TIBCO JasperReports, JasperServer.

PROJECTS

BMG Hack: Let's Get Visual- Civic Data Challenge *Python | Bokeh | Tableau | Altair* *Feb 2020*
- Explored and analyzed Bloomington city data about potholes, trails and storm water sewers.
- Created Bokeh, Altair and Tableau based interactive visualizations and presented them to Innovation department and residents.

Sentiment based Image Captioning *Python | tensorflow 2.x |* Numpy *| Google Colab | 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.

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.