

JOHN BATHAPPULLY

Using Data to Drive Impact

✉ Johnbathappully1998@gmail.com

☎ +91-9074173398

in [Linkedn](#)

🐙 [Github](#)

📍 Bangalore

🌐 [Youtube](#)

PROJECTS

Facial Analysis System

- Performed face **recognition**, **detection** and **tracking** with haar cascades, hog, lbph, and csrt tracker from a 50 k dataset
- Categorized **posture** and **emotions** with 73% accuracy and furnished **style** transfer of background/ **deep dream** on faces
- Generated 30% fake faces with **GAN**, to compensate for imbalance, and **compressed the segmented** images by vae

News Media Analysis System

- Probed areas of **expertise /bias** of journalists and articles **classified** to 5 topics on a 5 k imbalanced dataset
- Conducted **latent semantic/sentiment** analysis with an accuracy 90 % after removing the false news
- Grouped by **topic modeling** into 5 and **pos tagged** to identify grammatical errors in **summarized** and rephrased articles
- Employed **ner** with 83% accuracy to extract org/place names for news search optimization and **recommended** articles

Stock Market Prediction System

- Extracted/**scraped web** for **real-time** sentiment analysis of 5000 trending tweets/news headlines
- Projected bitcoin prices for next 30 days using **aws** and set up a **bot** that connects to **Zerodha** api for order placement.
- Simulated with ets, **arima**, varma, **garch**, **lstm-gru**, **nbeats** which were able to defeat naive forecast by more than 15%

Movie Recommendation system

- Devised hybrid model ensembling **content/collaborative**-based/**Rbm** models to reach a hit rate of 9% on 100 k records
- Enforced **stoplist**, eliminated 5% outlier users, and **Decomposed** to lower dimensions by svd
- Expanded by **scaling up** on 10m dataset with **Spark ALS/Sagemaker** and Calculated hitrate, coverage, diversity, error

Data engineering E-commerce system (AWS)

- Systematized a warehouse ie updates with orders and **operational/log analysis** system ie queries/alerts on 25% traffic
- Pictured features for **patterns** of order and automated **product recommendations** by 30 % via pyspark/sagemaker

PROFESSIONAL EXPERIENCE

AI Intern

CodeClause

📅 1st Feb, 2023 - current

- Generated additional** 2 x times images by **augmentation** and experimented for ideal neural network layers

Data Science Intern

The spark foundations

📅 5th Jan, 2022 - 1st Feb, 2023

- Applied **sentiment analysis** of stock news headlines and **forecasted 15 %** above random walk

Data science Intern

Oasis Infobyte

📅 2nd Dec, 2022 - 28th Dec, 2022

- Classified hate speech** with 94 % accuracy from tweets and **topic modeled** the offensive words for improving prediction

SKILLS

- Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Statsmodel, Beautifulsoup, Tensorflow, Nltk, Gensim, Spacy, Textblob, Opencv, Dlib, PySpark
- Data Science:** Web Scraping, Data Wrangling, feature engineering, Exploratory data analysis, Feature selection, Modelling, Deployment, Time Series, Recommender Systems
- Supervised Algorithms:** Linear, Logistic, Decision Tree, RandomForest, K-NN, Kernel SVM, Naive Bayes
- Unsupervised Algorithms:** KMeans, Hierarchical clustering, PCA, LDA
- Model selection & Boosting::** k-Fold, Grid Search, Xgboost, AdaBoost
- Evaluation metrics:** Adjusted R-Squared, Confusion matrix, Log-loss, Auc-Roc, gain-lift charts
- AI:** Natural Language Processing, Computer Vision
- Deep Learning:** Ann, Cnn, Rnn, Rbm, Autoencoder, Gan's, Transfer Learning, Hyperparameter Tuning
- Data Visualization:** Excel, Tableau
- Databases:** MySQL, MongoDB
- Aws Services:** kinesis, DynamoDB, EMR, RedShift, SageMaker, Lambda, OpenSearch, Glue, Athena, Kibana, Quicksight
- Spark:** Rdd operations, MLLib Modelling, streaming
- Hadoop:** Hdfs, Yarn, Sqoop, MapReduce, Flume, Pig, Hive, Hue, HBase
- MLops Life Cycle**

EDUCATION

B.Tech

Cochin University of Science and Technology 2017-2021
8.2/ 10

Awards

- ECSCA Branch Chair-Person**
- Approved **Phase1 funding** from Teqip for Hexapod crawler with Robotic arm

CERTIFICATIONS

- Data Science** Master's Simplilearn in collaboration with IBM
- Artificial Intelligence Engineer** Master's Simplilearn in collaboration with IBM
- Big Data Engineer** Master's Simplilearn in collaboration with IBM
- Machine Learning** by Stanford University
- Deep Learning** Specialisation by Deeplearning.ai