YASH PASAR

(315) 450 1430 | LinkedIn: linkedin.com/in/yashpasar/
Email: yspasar@syr.edu | GitHub: yashpasar | Portfolio: yashpasar

EDUCATION

Syracuse University, School of Information Studies, Syracuse, NY — *Aug 2018 - May 2020*
MS in Information Management, C.A.S in Data Science — CGPA: 3.67
Courses: Data Science, Big Data, Data Analytics, Natural Language Processing, Business Analytics, Database Admin Concepts & Management

University of Mumbai, Mumbai, India — *Aug 2014 - May 2018*
B.Tech. Computer Engineering — CGPA: 3.52
Courses: Algorithms, Data Structures, Databases, Operating Systems, Software Engineering, Data Mining

TECHNICAL SKILLS

Languages: Python, R, Spark SQL, SQL(T-SQL, SQL server, PostgreSQL), NoSQL(MongoDB, Cassandra, Neo4j), Hive, HTML, Java, C , C++
Libraries & Packages: TensorFlow, Keras, Scikit-Learn, NLTK, MLLib
Tools & Frameworks: Spark, Tableau, SAP(ECC, HANA, IQ), Splunk, Databricks, Hadoop, HBase, Amazon Redshift, Google Analytics
ML & Analytics: Classification, Regression, Clustering, Neural Networks, Bayesian Inferences, Time Series Analysis

PROFESSIONAL EXPERIENCE

Business Data Analyst Intern | Illumina, Inc., San Diego, CA — *May 2019 - Aug 2019*
- Leveraged Splunk and SQL to analyze the Illumina API call logs to understand end-user behaviors, requirements, pain points, monitor failures and diagnose root causes
- Designed Tableau dashboards to monitor key metrics of the API call logs in real time; reduced overall process time by 60%
- Worked with cross functional stakeholders to understand requirements, and used data-driven analytics to design and evaluate A/B Tests to maximize click rate and conversion rate

Business Data Analyst | iConsult Collaborative – Project Management Institute, Syracuse, NY — *Apr 2019 – May 2020*
- Analyzed web traffic using Python and Google Analytics from client website of 5K users to identify key KPIs (such as click rate, bounce rate, session details) for member acquisition & retention
- Conducted stakeholder interviews for requirement gathering and identifying important features
- Developed machine learning models (decision trees, neural nets and logistic regression) to predict individuals of a marketing campaign that are more likely to get converted into customers, with an accuracy of 92% and F-1 score of 93%

Graduate Teaching Assistant – Big Data Analytics | Syracuse University, Syracuse, NY — *Jan 2020 – May 2020*
- Mentored students in machine learning concepts (classification, regression, trees, deep neural networks and cross-validation)
- Guided students on handling large datasets with Apache Spark, MapReduce and Databricks/AWS and designed projects to test student understanding of ML concepts and their Python implementations
- Conducted research in Big Data technologies, such as Apache Spark & NoSQL, to design a new curriculum for future cohorts

Business Technology Analyst | iConsult Collaborative – One Planet Edu. Network, Syracuse, NY — *Aug 2018 - Apr 2019*
- Optimized internal processes by planning and analyzing complex, strategic business problems solved using automated systems
- Interfaced with business stakeholders and IT teams to elicit, analyze, validate and document business requirements (BRD)
- Developed user stories to glean user requirements using data analysis/feedback for gap analysis and business process mapping

PROJECTS

Walmart Store Sales forecasting | Python, Flask, Heroku — *Aug 2019 - Dec 2019*
- Performed exploratory statistical data analysis for 400K records to recognize the key factors that boost the weekly sales
- Developed and fine-tuned KNN, Linear Regression, SVM and Ensemble models (such as random forest), with accuracy of 95%
- Developed a web app to perform online prediction of weekly sales with Dash & deployed it on Heroku [webapp]

Loan Defaulter Prediction | Python, Spark, Databricks — *Aug 2019 - Dec 2019*
- Performed ETL on 4M financial records using Spark SQL to discover patterns and trends of loan defaulters and leveraged demographic (like age, occupation) and financial (like FICO, income) to predict probability of defaulting on loan payment
- Developed four models (logistic regression, SVM, Naïve-Bayes and random forest) and handled class imbalance by under sampling the data to reduce bias
- Used MLLib to compare models and tuned hyperparameters to achieve the highest F1-score of 0.89 with Random Forest

NYSE Stock Trading System | MySQL, Flask, MS Visio — *Jan 2019 - Apr 2019*
- Designed and developed a database management system to store, manipulate, and search stocks for an investor's portfolio and translated staff designed ER model into a relational schema implemented with MySQL and Flask
- Added functionalities like insert, delete, update, triggers and generating reports; analysed query plans and optimized the queries

Brain Tumor Detection and Classification | MATLAB — *May 2017 - May 2018*
- Designed pipeline consisting of image segmentation, feature extraction, and classification to detect tumors in images
- Experimented with various features (segmentation, edges, clustering) and models (CNN, SVM) to achieve 92% accuracy
- Packaged the classification pipeline in an application, presently in use at KJ Somaiya Hospital & Research center
- Published and presented a survey research paper in INDIACom'18, IEEE conference [research paper]