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 — *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 — *May 2018*
B.Tech. Computer Engineering — CGPA**:** 3.52

TECHNICAL SKILLS

Languages: Python, R, Spark SQL, SQL(T-SQL, SQL server, PostgreSQL), NoSQL(MongoDB), BigQuery, Git
Tools & Frameworks: Spark, Tableau, SAP HANA, Dash, Splunk, Heroku, Azure Databricks, Hadoop, Amazon Redshift, Google Analytics
Libraries & Packages: Scikit-Learn, MLlib, NLTK ,Word2Vec , Keras, TensorFlow, Matplotlib, Seaborn,
ML & Analytics: Classification, Regression, Clustering, Neural Network, Deep Learning, Bayesian Inferences, Time Series Analysis

PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant – Big Data Analytics | Syracuse University, Syracuse, NY — *Jan 2020 - Present*
- Assisted up to 6 students daily with solving problems in MapReduce & Apache Spark for on-premise and cloud (Databricks/AWS)
- Conducted research in Big Data technologies like Apache Spark & NoSQL to design a new curriculum for future cohort
- Designed assignments to test ML skills for large-scale datasets using Spark SQL & MLlib for a batch of 100 students

Business Data Analyst | iConsult Collaborative – Project Management Institute, Syracuse, NY — *Apr 2019 - Present*
- Analyzed web traffic using Google Analytics from client website of 5K users to identify KPIs for member acquisition & retention
- Designed SEO strategies, website implementation plan and web platform comparison reports to increase members by 10%
- Conducted stakeholder interviews for requirement gathering on existing system infrastructure and marketing strategy
- Developing an improved web platform using WordPress for better data management and provide support for maintenance

Business Data Analyst Intern | Illumina, Inc., San Diego, CA — *May 2019 - Aug 2019*
- Analyzed large datasets using complex SQL queries to solve issues faced on the eCommerce website for seamless user experience
- Automated manual processes on Splunk & Tableau by creating 12 varied interactive dashboards to reduce process time by 60%
- Delivered business reviews to senior executives about progress and roadblocks to provide a unified solution for 5 teams
- Developed business requirement documents for orders and quote pricing on SAP from various stakeholders

Business Technology Analyst | iConsult Collaborative – One Planet Edu. Network, Syracuse, NY — *Aug 2018 - Apr 2019*
- Gathered stakeholder input and feedback through interviews and surveys to develop DFD, ERD & UML diagrams
- Identified process boundaries and determined opportunities to automate processes and functions
- Analyzed historical data to generate projections, developed key business insights and addressed data discrepancies

PROJECTS

Brain Tumor Detection and Classification | MATLAB
- Built pipeline to detect & classify tumors containing Image Pre-Processing, Segmentation, Feature extraction & Classification
- Extracted features to build a hyper-parameter tuned CNN model to isolate the tumor with an accuracy of 92%
- Packaged the classification pipeline in an application built with MATLAB employed by KJ Somaiya Hospital & Research center
- Published & Presented research paper in INDIACom'18 [research paper]

Walmart Store Sales forecasting | Python, Flask, Heroku
- Performed EDA & data wrangling for 400K records to understand the key factors that boost the weekly sales using Python
- Built and fine-tuned 4 different machine learning models to achieve an RMSE of 5%
- Developed a web app to perform online prediction of weekly sales for stores with Dash & deployed it on Heroku [webapp]

Big Data, Big Bank | Python, Spark, Databricks
- Analyzed & performed ETL using Spark SQL on 4M financial records to discover patterns and trends of loan defaulters
- Handled class imbalance by under sampling the data to reduce bias & built multiple ML models using MLLib
- Tuned hyper parameters to achieve the highest F1-score of 0.76 with Random Forest

NYSE Stock Trading System | MySQL, MS Access, MS Visio
- Created a database management system to simplify searching, buying & managing of stocks through individual user portfolio
- Designed ERD and relational schema for the system and designed application using MS Access and SQL
- Created tables and databases in MS SQL and added functionalities like insert, delete, update, triggers and generating reports

CERTIFICATIONS

- **Coursera (Mar 2020):** Machine Learning by Stanford University
- **Google (Mar 2020 – Mar 2023):** Advanced Google Analytics
- **Lynda (Feb 2019):** Python for Data Science Essential Training