Saurabh Shrinivas Maydeo

Software Development Engineer

SaurabhMaydeo@gmail.com Saurabh Shrinivas Maydeo

Research Interest

Machine Learning, Computer Vision, Medical Data Analysis, Healthcare

Education

Master of Science in Computer Science **Jan 2020 - Dec 2021**
Georgia State University, GPA: 3.83/4.00 *Atlanta, GA*

Bachelor of Engineering in Information Technology **Aug 2015 - May 2018**
University of Mumbai, GPA: 08.74/10.00 *Mumbai, India*

Research Experience

Data Mining Lab, Georgia State University **May 2020 – Aug 2020**
Graduate Research Assistant *Atalanta, Georgia*

- Implemented local spatiotemporal outlier detection algorithms to detect outliers from magnetogram files representing solar-astroinformatic data
- Utilized K-means++ algorithm to form clusters based on movement characteristics and performed dissimilarity comparison to quantify the similarity between query and template trajectory segments. Validated the method by carrying out 3 case studies
- Published paper titled "Local Outlier Detection for Multi-type Spatio-temporal Trajectories" in the 6th Special Session on Intelligent Data Mining at the 2020 IEEE International Conference on Big Data
- Specific contributions: Handled vast numbers of magnetogram files, performed data preprocessing, integration, and exploratory data analysis, and assisted in performing experiments

Publications

- **"Local Outlier Detection for Multi-type Spatio-temporal Trajectories"** in 6th Special Session on Intelligent Data Mining in 2020 IEEE International Conference on Big Data.
- **"Fake Currency Detection and its Associated Classification"** in IJIRCCE, Volume 5, Issue 12, ISSN: 17554 – 17561.
- **"Identification System based on Biometric Concept Using Human Eye"** in IJIRCCE, Volume 5, Issue 12, ISSN: 17507 – 17512.

Industry Experience

Amazon **Feb 2022 – Present**
Software Development Engineer *Seattle, Washington*

- Implemented a wrapper library around Kafka consumer for a streaming service that different consumer applications in the ad delivery system can use to listen to events on specific topics such as ad products, targeting, etc.
- Designed and developed a near real-time service using Kafka to sync publisher ad products created on the supply side with the demand side of advertising.
- Designed business logic and DAO layer for supply-side services at Amazon Advertising to persistently store objects based on dynamically generated schema.
- Developed and deployed an Event Producer library that simplifies event publication for clients by abstracting topic management and Kafka producer configurations.
- Created a producer-lambda that auto-responds to storage updates, publishing events to relevant topics/partitions via the Event Producer library.
- Developed Integration test framework using TestNG and Hydra for CRUD operations on publisher-specific objects. Collaborated with principal engineers for code reviews, ensuring high-quality standards.
- Conducted load and stress tests with Hydra, TestNG, and Cloudwatch to identify vulnerabilities and areas for service improvement. Documented results, findings, and detailed analysis.
- Resolved production issues by thorough and meticulous troubleshooting and investigation.

Truist **June 2021 – Aug 2021**

Data Science Accelerator Intern *Atalanta, Georgia*

- Leveraged machine learning algorithms to extract actionable insights from the Retail Deposit Attrition Model, identifying customers likely to close deposit accounts with Truist within 90 days. Analyzed key attrition factors to inform targeted retention strategies. Presented findings through data visualization techniques. Predicted a 2.2% attrition rate, potentially saving approximately $7.5 million in costs.
- Developed hackathon project analyzing call center data, identifying trends and insights in customer care communication. Utilized advanced visualization techniques for presenting key findings and compelling narrative.

Teaching Experience

Georgia State University - CS 8850: Advanced Machine Learning **Sept 2020 – Dec 2021**

Graduate Teaching Assistant *Atalanta, Georgia*

- Provided support to Dr. Sergey Plis in teaching Advanced Machine Learning for three consecutive semesters.
- Evaluated and graded assignments, projects, and papers
- Conducted regular doubt-solving sessions to assist students in comprehending complex concepts and problem-solving strategies.
- Facilitated effective communication between students and the course instructor, helping to address individual learning needs.

Georgia State University - CSC 3210: Computer Organization Programming **Jan 2020 – May 2020**

Graduate Teaching Assistant *Atalanta, Georgia*

- Hosted 1 hour-long weekly as a course instructor for the Computer Organization programming course under Dr. MD Zulkar Nine.
- Evaluated and graded assignments and conducted regular office hours to assist students with any questions

Technical Skills

Languages: Java, Python, GraphQL, CQL, Javascript, PHP, C, C++, HTML, SQL, Typescript

Technologies / Libraries / Data Formats: Pandas, NumPy, sci-kit-learn, Keras, LaTeX, Matplotlib, AWS, S3, MSK, RDS, Apache Kafka, DynamoDB, Lambda, Cloudwatch, Neo4j, Node.js, Spring Framework, Hydra, Git, Docker, Jenkins, PostgreSQL, MongoDB, React, TestNG, JUnit, Mockito, Bootstrap, Flask, jQuery, Apache Spark, Kubernetes, Linux, Unix, etcd, Nginx, JSON, XML

Coursework

Advanced Machine Learning, Deep Learning, Machine Learning, Fundamentals of Data Science, Design and Analysis of Algorithms

Projects

Pulmonary Fibrosis Progression - Predicting lung function decline | *Python, sklearn, keras, Tensorflow, NumPy*

- Developed a predictive model for pulmonary fibrosis to forecast the severity of a patient's lung function decline using CT scans of their lungs and baseline forced vital capacity (FVC) measurements.
- Conducted exploratory data analysis and preprocessing to prepare the dataset for modeling.
- Experimented with two different models, EfficientNet B5, and Quantile Regression Dense Neural Network, to predict the decline in lung function and blended the outputs of both models to leverage the diversity in their predictions.

Breast Cancer Classification | *Python, keras, TensorFlow*

- Developed ML model to classify patients into malignant and benign breast cancer to help in early diagnosis of breast cancer, which can improve the chances of patients getting good treatment from the beginning and thus increasing chances of their survival.
- Developed and trained several machine learning models – Random Forest, Logistic Regression, SVM, kNN, Naïve Bayes classifier and chose Random Forest classifier as the best model with 96.50% accuracy and 95.47% recall.

Classification of tweets into disaster and non-disaster tweets: | *Python, Sklearn, TensorFlow*

- Built ML model to classify tweets into disastrous and non-disastrous tweets.
- Data preprocessing (handling misspelled words, URLs, hashtags, etc.)

- Developed and trained several machine learning models - SVM, XGBoost, LSTM, and BERT and chose BERT as the best model and got an 80.90% F1 score on the test dataset.

Cloud based File Storage System | *Java, HTML, CSS, JavaScript, AWS S3, AWS RDS, React*

- Created a cloud-based file storage system supporting various file formats, such as images, videos, and text, inspired by Google Drive, featuring user-friendly APIs like ListDirectories, ListFiles, GetFile, UploadFile, DeleteFile, CreateFolder, and DeleteFolder, streamlining file manipulation within the storage system.
- Utilized AWS S3's BLOB storage to optimize the storage and management of files and directories, ensuring efficient and reliable data handling.
- Leveraged AWS RDS with PostgreSQL to efficiently store and retrieve metadata associated with files and folders, enhancing data organization and accessibility.

Certifications

- **"Deep Learning",** a 5-course specialization by deeplearning.ai(Andrew Ng's) on Coursera. Specialization Certificate earned on August 3, 2018
  https://www.coursera.org/account/accomplishments/specialization/certificate/T5XDCQFJ7V78
- **"Python for Everybody",** a 5-course specialization by University of Michigan on Coursera. Specialization Certificate earned on April 11, 2019 https://www.coursera.org/account/accomplishments/specialization/RJ976MZBUEJX
- **"Machine Learning",** by Stanford University on Coursera. Certificate earned on Wednesday, June 20, 2018, 11:41 AM
  https://www.coursera.org/account/accomplishments/certificate/B52F6Z6B226N