Sayan Biswas

Boston, MA 02120 | sayanbiswas07@gmail.com | +1-(617)-516-4883
linkedin.com/in/sayanbiswas59 | github.com/sayanbiswas59

EDUCATION

Northeastern University, Boston, MA **May 2021**
Master of Science in Data Science **GPA: 3.94/4**
Related Courses: Data Management and Processing, Algorithms, Supervised Machine Learning, Information Retrieval, Causal Modeling in Machine Learning, Unsupervised Machine Learning and Data Mining, NLP, DBMS
Activities: Vice President of Husky Data Club, Global Student Mentor for 100+ incoming students

Visvesvaraya Technological University, Bangalore, India **Jun 2013**
Bachelor of Engineering in Electronics and Communication **GPA:83/100**

PROFESSIONAL EXPERIENCE

Danfoss Power Solutions, Cambridge, MA
Data Science Intern **Sep 2020 – Dec 2020**

- Designed a robust **machine learning framework** around Danfoss's proprietary simulation environment, allowing for rapid and safe prototyping of autonomous vehicle functionality
- Engaged, collected, and labeled images leveraging the framework to train and simulate a deep learning-based use case
- Researched and prototyped smart lane-keeping by training a neural network architecture utilizing **CNN** to take full steering control of a vehicle within simulation environment, accurately replicating human operators by **88**%
- Devised a novel **anomaly detection method** that employs **k-means clustering** and **Random Forest** to detect & flag anomalies in real-time in production system, as well as **Shapley values** to increase model interpretability
- Implemented a deployment-ready **sales forecasting model** leveraging **Azure ML,** achieving an RMSE of **$15**, which was then incorporated with **Power BI** to aid critical business decision-making process
- Presented findings & bottlenecks to external teams and management, and documented detailed reports on its approach

Tally Solutions Pvt. Ltd., Bangalore, India
Senior Software Engineer, Data Science **Aug 2016 – Dec 2018**

- Led a high performing analytics team within **R&D** in an **Agile** environment, offering solutions & insights for complex & unstructured business challenges using data-driven techniques
- Formulated, modeled & analyzed key metrics to segment customers and forecast customer enrollments and sales patterns via ARIMA model and SVM Regression
- Collaborated to develop a robust **ETL** pipeline that unifies and transforms Sales, Marketing & CRM data stored leveraging **AWS** managed services, resulting in a **45%** increase in data analysis performance
- Developed a **customer churn prediction model** with **87%** accuracy that helped the organization send timely emails to retain customer subscriptions and increased **ROI** by **$2800** every month
- Proactively modified incoming service **email-classifier** with **word-to-vector model**, resulting in 30% reduction in workload for Customer Care representatives over the previous algorithm

Software Engineer **Oct 2013 – Jul 2016**

- Designed a normalized (**3NF**) physical data model as a Master database to enable single point reference to **ERP** data
- Performed data processing & calculated various KPIs, developed interactive dashboards and reporting workbooks utilizing **Tableau** and **MySQL**, resulting in a ~40% faster assessment of the **KPIs**
- Designed **Python-based** robust automation framework to automate manual testing suite of **RESTful API's** for data management, licensing, tax filing operations, reducing manual testing effort by **60%**
- Improved test processes by developing automation tool while also integrating **Git** & **Jenkins**, thus eliminating manual effort in testing of daily builds of **ERP** software product by **80%**

TECHNICAL SKILLS

Languages:	Python, R, C#, C++, SQL, XML, JSON
Databases:	MySQL, PostgreSQL, SQLite3, MongoDB
Libraries & Framework:	Pandas, NumPy, Scikit-Learn, SciPy, TensorFlow-2.0, PyTorch, Flask, OpenCV, PySpark, Plotly, tidyverse
Software/Cloud:	Jupyter, RStudio, Visual Studio Code, Tableau, AWS, Azure, GCP, Jenkins, JIRA, Git, Docker, Kubernetes
Machine Learning:	Regression, Classification, Clustering, Dimensionality Reduction, Random Forest, XGBoost, Neural Nets, Deep Learning (CNN, RNN, LSTM), Natural Language Processing(BERT), Decision Trees, Naïve Bayes

ACADEMIC PROJECTS

Movie Recommendation System: [Python]

- Developed and deployed via **Flask** a recommendation system that uses **Collaborative Filtering** with **Matrix Factorization** to generate recommendations based on 2.5 GB of user submitted movie ratings data, with a 92.5 % predictive accuracy

Loan Prediction Tool: [Python, Flask, AWS, Tableau]

- Analyzed HMDA dataset of 15 Million records, 99 features and accounted for class imbalance using **SMOTE**, applied **LASSO** for feature selection and achieved an accuracy of 87% by iterative tuning of **Random Forest** algorithm

Abnormality detection in Musculoskeletal Radiographs (MURA): [Python, TensorFlow]

- Detected and localized abnormalities in X-Ray images of different body parts by training **169-layer DenseNet** with **transfer learning**; thereby achieving a recall of 87% on test set