Swastik Biswas

Email: swastikbiswas2016@gmail.com Mobile: +91-905-1652-880

Self-motivated and a results-oriented Data/ML engineer with 2+ years of professional experience. My interest lies in learning how we use computer vision and image processing to impact people's lives. Real-world scene understanding, activity recognition, anomaly event detection, extreme event forecasting in a computer vision setting are some of the things that are always exciting to me. I am equally enthusiastic about discovering how we can combine computer vision with natural language processing for innovating approaches in which we can extend human interaction with computers.

Education

- **Jadavpur University** — Kolkata, India
  Master of Engineering - Computer Science and Engineering; CGPA: 8.11 — *August 2017 - June 2019*
  Thesis: *"Volumetric Brain MRI Segmentation using Entropy based Fuzzy clustering algorithm"*
  Advisor: *Prof. Dr. Jamuna Kanta Sing*

- **Visvesvaraya Technological University (VTU)** — Bangalore, India
  Bachelor of Engineering - Computer Science and Engineering; CGPA: 7.31 — *August 2012 - June 2016*

Research Experience

Academic projects

- **Entropy based Fuzzy clustering algorithm for MR image segmentation** — Jadavpur University
  Master's Dissertation — *Sep. 2018 - Mar. 2019*
  Used Shannon entropy to modify the Fuzzy C-Means algorithm for segmenting volumetric brain MR images, model was evaluated on the Brainweb dataset and IBSR dataset before utilizing for the real-patient image volumes.

- **Gaussian Energy-based function for detection of moving objects in videos** — Jadavpur University
  Research project — *Feb - May. 2019*
  Utilized Gaussian distribution to modify energy-based functions in images for motion detection of objects to identify accidents in a real-time video. A dataset consisting of 3320 contain crash frames and 6180 normal frames, ranging from a duration of 3-5min were curated from online sources since there was no public database for vehicle accidents or crash detection. Accuracy achieved during training phase was around 84%.

- **Deep Learning through Image Analysis of real-time videos** — VTU
  BE project — *Feb. - May.- 2016*
  Classified entities from real-time videos and presented a comparative study on various models. Image classification was compared on deep belief networks, deep feedforward neural networks, and convolutional neural networks.

Industrial projects

- **Fuzzy graph networks for building contextualized knowledge graphs** — BRIDGEi2i Analytics Solutions
  Research project — *May 2021 - Present*
  Built a labeled property graph that can be used for representing contextual information across various documents. Fuzzy searching algorithms are used along with graph neural networks for achieving better properties on node values. The model is currently used for verification across different data and in different proof-of-concept(POCs).

- **Automated Driver Assistance Systems** — ITC Infotech
  Internship — *Jul. - Aug. 2015*
  Developed a real-time video synthesis application on the concept of Automated Driver Assistance Systems for extracting textual information from real-time videos. A CNN model was trained over a custom dataset after evaluating CIFAR-10 and CIFAR-100, for classifying objects of interest. Tesseract OCR was used for extracting any text present in the videos. Accuracy achieved during training and validation phase was around 94% and 89% respectively.

Publications & Presentations

IEEE Conference Papers (Peer-Reviewed)

- S. Biswas, N. Mahata and J. K. Sing, *"A New Entropy Based Fuzzy Clustering Algorithm for Volumetric Noisy Brain MR Image Segmentation,"* 2019 Fifteenth International Conference on Information Processing (ICINPRO), 2019, pp. 1-4.

Professional Experience

Lead Data/ML Engineer - BRIDGEi2i Analytics Solutions — ***Jul. 2019 - Present***

- Data Federated Services Platform — Dec. 2020 - Present
  Building data federated platform for one of the largest American multinational Platform-as-a-Service companies with the scope of providing an integrated customer recommendation experience. Along with GloVe vectors for similarity, BERT fine-tuned on a custom dataset is used as a recommendation model with a validation accuracy of approximately 91%. Currently, the production platform handles data at a scale of over 10TB daily with the scope of increase in the future.

- Market share prediction (POC) Oct. - Dec. 2019
  Stochastic differential equation with Gaussian approximation was used for POC focused on understanding the market trend using the sales information for one of the largest British multinational beverage alcohol companies. By performing the simulations for each subsequent point and with a given threshold value, the range of the next value was found. The process is used for identifying anomalous data points within a given window frame. The use case originally had 34 key performance indicators(KPIs) before trimming to 28 KPIs and involved a monthly data refresh of approximately 3GB.
- Data modelling and analysis Aug. - Dec. 2019
  Built ingestion pipelines, prepared denormalized data, and utilized residual sum of squares to detect change points for identifying a level shift in the time series data (in the HR domain) for one world's leading professional services firms. Also, KNN causal estimation was used for determining the causal strength of KPIs after using partial correlation to define independence between the variables. The use case involved had 22 KPIs along with a monthly data refresh of roughly 25GB.
- Logistics Demand analysis Jan. - Jul. 2020
  Used the exponential moving average for detecting anomalies in the historical logistics data for a leading Indian consumer goods company. Furthermore, relationships across various KPIs were determined using mutual information and the chi-square test across the variables. Originally the project started with 7 KPIs and was later increased to 15 KPIs with a bi-weekly data refresh of around 2GB.
- Data Ingestion and Topic modelling Jul. - Dec. 2019
  Along with building pipelines for data ingestion in Talend and preparing the data model, I was involved in performing topic modeling and sentiment analysis on text data of call transcripts for one of the largest American telecommunications organizations. The pipelines involved real-time model scoring with a daily data refresh of about 3-5TB.

Professional Certifications AND Achievements

- **Natural Language Processing Specialization by DeepLearning.AI** 20 Feb, 2021
  Coursera
- **Deep Learning Specialization by DeepLearning.AI** 02 Aug, 2021
  Coursera
- **Preparing for Google Cloud Certification: Cloud Data Engineer** 29 Apr, 2021
  Coursera
- **Awards from Technology Labs**
  BRIDGEi2i Analytics Solutions
  - **Individual award for Above and Beyond performance in 4th Quarter of financial year 2020**: 23 Feb, 2021
  - **Team Award for outstanding performance in 2nd Quarter of financial year 2020**: 13 Aug, 2020
  - **Team Award for outstanding performance in 2nd Quarter of financial year 2019**: 10 Sep, 2019

Scholarships & Grants

- **AICTE PG Scholarship** 2017 - 19
  Post-Graduation scholarship provided by AICTE
- **SERB, DST. Govt. of India** 2018 - 19
  Research grant by Dept. of Science & Technology

Technical Skills

- **Languages**: Python, Java, C, C++, Go, JavaScript, HTML, CSS
- **Databases**: MySQL, PostgreSQL, Neo4j, Redis, Hadoop, Hive
- **Frameworks**: Flask, FastAPI, TensorFlow, PyTorch, Apache Solr, Elasticsearch, Spring, Express JS, Bootstrap
- **Tools**: Git, Docker, Kubernetes, Talend, Jenkins, Jira

Activities & Interest

Literature, Art, Coding, Cycling, Travelling