Shubham Mawa
shubhammawa.github.io
Email : shubhammawa1998@gmail.com
Mobile : +91-8699519186

Education

- **Indian Institute of Technology, Kharagpur** — Kharagpur, West Bengal
  M.Tech + B.Tech Dual Degree in Industrial and Systems Engineering; GPA: 8.31/10 — *July. 2016 – April. 2021*
  Minor in Mathematics and Computing
- **G.M.S.S.S, Manimajra** — Chandigarh, India
  Central Board of Secondary Education; Standard 12th; Percentage: 85.6 — *2016*
- **Little Flower Convent School** — Panchkula, India
  Indian Certificate of Secondary Education; Standard 10th; Percentage: 96.6 — *2014*

Internships

- **AWL** — Sapporo, Hokkaido, Japan
  Artificial Intelligence Research Internship — *Apr 2020 - Aug 2020*
  - **Semi-Supervised Learning: MixMatch**:
    - Implemented MixMatch, a semi-supervised learning method to leverage unlabelled data to train classification models.
    - Created a pipeline for creating augmentations for unlabeled data, generating guessed labels and sharpening.
    - Designed experiments for performance comparison of MixMatch with fully supervised baselines and performed an ablation study. Implemented variants of Mixup algorithm and analyzed loss function behaviour.
    - Achieved an F1-score within 3% of a fully supervised model using only one-tenth of the total labels.
    - Studied research papers on consistency regularization, entropy minimization, mixup and pseudolabeling.
  - **Age and Gender Estimation**:
    - Trained a gender classification model on face images using a VGG16 model and achieved an F1-score of 0.93.
    - Studied key dataset attributes for age estimation from face images like pose, lighting, expressions, age range etc and worked on improving dataset quality and model robustness.
    - Used two methods, Rank Consistent Ordinal Regression and Two-point Representation for estimating age. Used a classifier importance parameter to penalize large classification inconsistencies.
    - Designed experiments to assess model performance with augmentation pipeline and ranking consistency.
  - **Data Labelling**:
    - Analyzed model performance decline in age and gender estimation due to use of masks under Covid-19 circumstances and identified need for a new labelled face dataset with masks.
    - Used pseudo-labeling, a self-learning method to filter high confidence predictions and label the dataset iteratively.
    - Performed outlier detection and used sampling methods for manual checking and correction.
- **PredictivEye** — Toronto, Ontario, Canada
  Data Science Internship — *Dec 2019 - Jan 2020*
  - **Recommender Systems**:
    - Developed product recommendation systems using collaborative filtering and content-based method for a customer-analytics platform. Incorporated user journey information and tracked browsing behaviour.
    - Developed an inventory analytics and visualization tool with insights like most-purchased, most-viewed items etc.

Coursework

- **Computer Science:** Artificial Intelligence, Deep Learning, Machine Learning, Natural Language Processing, Data Structures and Algorithms, Object Oriented System Design, Symbolic Logic
- **Mathematics:** Linear Algebra, Probability and Statistics, Regression and Time Series, Stochastic Processes, Statistical Decision Modelling, Applied Multivariate Statistics
- **Industrial Engineering:** Operations Research, Optimisation and Heuristic Methods, Simulation, Quality Design and Control, Safety Analytics, Game Theory, Supply Chain, Production Planning, Inventory Systems
- **Miscellaneous:** Recommender Systems, Information Systems, Cognitive Information Processing, Psychology of Learning, Product Development, Economics, Electrical Technology, Basic Electronics

Projects

- **Text Classification using CNN** — Prof. Sudeshna Sarkar
  Term Project: Natural Language Processing — *Oct 2018 - Jan 2019*
  - **Automatic Event Extraction from News documents**:
    - Classified documents into predefined event types like 'Heat Wave', 'Earthquake', 'Storm', 'Cyclone' etc, where events depict occurrence of natural or man-made disasters.
    - Generated word vectors using SkipGram and C-BOW models and paragraph vectors using the Fasttext library.
    - Recognized words which triggered the detection of particular events using a combination of CNN and Bi-LSTM
  - **Sentiment Analysis**:
    - Classified reviews from websites like Amazon, IMDB, Yelp into positive and negative sentiments.
    - Implemented a single-channel Kim-CNN text classification model with modifications in filter dimensions to handle a sentence feature map. Used static as well as fine-tuned word-vectors as input.
- **Model Fitting and Adequacy Checking** — Prof. Buddhananda Banerjee
  Term Project: Regression and Time Series Analysis — *Nov 2018*
  - Performed regression analysis on a bicycle sales dataset. Plotted feature correlation values.
  - Calculated and compared adjusted R-squared values after feature selection.
  - Fitted an ARIMA model to capture trend and seasonality information and compared with regression model.
- **Queuing System Simulation of Railway Counters using ARENA Software** — Prof. Goutam Sen
  Term Project: Simulation Lab — *Jan 2019 - Apr 2019*
  - Identified important process parameters and variables and collected data like inter-arrival time, service time.
  - Performed input analysis to find the best fitting distributions for data collected for the different variables.
  - Built a model in ARENA considering different resources, constraints, process parameters. Compared simulation and theoretical outputs. Suggested changes to improve performance and efficiency.
- **Rash Driving Detection** — Prof. Goutam Sen
  Term Project: Work System Design Lab — *Jan 2018 - Apr 2018*
  - Collected data such as velocity, acceleration, steering angle etc on a bicycle using SensorLab app on Android.
  - Analyzed and plotted data to verify the reliability of data obtained from sensors.
  - Used machine learning models to classify the driving behaviour of a driver as normal or rash.
- **Unequal Area Facility Layout Modelling** — Prof. Manoj Kumar Tiwari
  Term Project: Optimization and Heuristic Methods — *Jan 2019 - Apr 2019*
  - Calculated dimensions and located rectangular facilities in an unlimited floor space, without overlap, while minimizing the sum of distances among facilities weighted by "material-handling" flows.
  - Used Genetic Algorithms to model and solve the problem. Studied parameters like facility-type, aspect ratios, material flow, and GA parameters like crossover probability, mutation probability etc.

Competitions

- **NetApp Data Challenge, Kshitij, IIT Kharagpur:** Secured 5th position overall and 3rd in model performance in a text classification problem among 150 teams.
- **Intelligent Ground Vehicle Challenge, Oakland University, Michigan, USA:** Represented IIT Kharagpur at 26th IGVC as a member of Autonomous Ground Vehicle Research Group and secured 2nd position in AutoNav Challenge. Led the mechanical team of the research group.
- **Strategy Storm, International Social Business Case Competition, IIT Guwahati:** Secured 3rd position among 1200 teams. Provided digital solutions for scalability and customer reach.

Skills and Expertise

Languages: Python, C, C++, SQL, Java **Frameworks**: Pytorch, Tensorflow, Sklearn, Pandas
Softwares: Git, CPLEX, MySQL, Ansys, Solidworks, Excel, Powerpoint