Abhishek Niranjan

abhishek.niranjan.iitkgp@gmail.com | abhishek-niranjan.github.io | https://in.linkedin.com/in/abhishekniranjan | +91-8513902048

ACADEMIC QUALIFICATIONS

Degree/Certification	Institution	CGPA / %	Year of Completion
B.Tech + M.Tech, Computer Science and Engineering	Indian Institute of Technology, Kharagpur	7.8/10	2018*
AISSCE(CBSE) (XII Grade)	Kendriya Vidyalaya No. 2, AFS Jodhpur	93.6%	2012
AISSE (CBSE) (X Grade)	Kendriya Vidyalaya No.1, Udhampur	9.8/10	2010

- **Joint Entrance Examination - Advanced | 2013:** Stood amongst ***top 0.5% out of 1.4 million students*** in JEE-Advanced 2013 and got shortlisted for admission in ***Indian Institute of Science, Bangalore***, the finest science institution in India.

INTERNSHIPS

Data Science Intern | Amplus Solar | May '17 - July '17

- Devised a novel way of computing soiling rate in Photovoltaic plants experiencing similar climatic conditions from limited data.
- Engineered a forecasting module to predict hourly active power generation by a PV plant using **gradient boosted trees,** achieving a correlation of 0.97, enhanced the feature set by utilising **OpenWeatherMap API.**
- Automated the generation of daily visualisation reports of PV Plants portfolio in **Power BI by connecting MySQL server** hosted on AWS EC2 instance.
- Built a GUI application using **PyQt4** to facilitate the data retrieval from multiple PV plants' dashboards hosted on several sites.

Data Science Intern | Bidgely Technologies | May '16 - July '16

- Studied the low-resolution electrical energy consumption data of US consumers to analyse the pattern in vacation periods.
- Implemented a **sliding window algorithm** to predict the vacation instances of the residents using **MATLAB** with **precision >95% and recall >70%**, which got incorporated into production.
- Worked on the **energy disaggregation** module to compute refrigerator consumption from low resolution, raw electrical energy consumption, data which was **pushed into** disaggregation pipeline in **production**.
- Received deep appreciation for projects completed and got **recommended for full-time position** by Head, Data Scientist.

Full Stack Developer | Outsy | May '15 - July '15

- Developed a polling feature for real-time meetings in **Django Web Framework** which was incorporated in the android app.
- Extracted artists' names from 15,000 Facebook textual posts using **Stanford POSTagger after NLTK** assisted pre-processing
- Generated Artists database by drawing out information using **Wikipedia API, Youtube API, and SoundCloud API.**

PUBLICATION

- **Joint Conference on Digital Libraries | Toronto, Canada | Jun '17:** Citation sentence reuse behaviour of scientists: A case study on massive bibliographic text dataset of computer science.

PROJECTS

Copying Citation Contexts : Natural Language Processing

- Analysed a massive dataset comprising of nearly 1.5 million computer science articles and more than 26 million citation contexts to understand the behaviour of "Copying Citation Contexts" amongst the researchers.
- Conducted experiments involving **cosine similarity measure** to reveal the copying pattern in 24 fields of computer science.
- Examined the variation, of copying phenomena in the most cited research papers, in subsequent 5 years of their publication.

Scientific Document Clustering Based on Key-phrases and Relations: Natural Language Processing

- **Extracted key-phrases and relations** from citation-contexts from research articles in NLP as Master's Thesis Project with the aim to create DBPedia RelFinder like relation extractor tool for research community in academia.
- Learned **vector embeddings** of research articles' bibliographic text**(doc2vec)** to create feature set for unsupervised clustering.
- Extended the feature set by augmenting **weighted link-cardinality** for each of the tagged research article to produce a supervised model for tagging research articles on the basis of Tasks solved by article.

Competitive Strength Prediction of ATM Vendors in California : Data Analytics

- Assessed the competitive strength of 3 major ATM vendors using spatial ATM network and the demographics of California.
- Visualised feature importance using **Tableau** and clustered the ATM locations by utilising **k-means approach**.
- Reverse engineered the feature weights by applying **Random Forest Classifier** on the clustered data for each county.
- Built a county specific **linear regressor** to model the revenue of each ATM. Enhanced the whole California state linear model by adding county-specific linear model to estimate the annual revenue generation of each ATM location.

Sign Language Translation Through Sensory Gloves : Machine Learning

- Worked in a team to translate American sign language gestures to text using flex sensing gloves with positional sensors.
- Compared different classifiers on input data from sensory gloves to train the alpha-numerical character recognition algorithm.

- Built a statistical language model based on stochastic grammar to recognise meaningful words from a stream of characters.

Topic Drift in Comments Section of Articles : Natural Language Processing

- Studied the phenomena of Topic Drifting in the comments section of an online article by identifying the comment which deviates - from the topics touched by the article or discussion in the prior comments.
- Applied **LDA (Latent Dirichlet Allocation) Model** to generate the topics associated with the content of the article. Thereby, tagging article sentences to the corresponding topic.
- Employed **cosine similarity metric** to map comments to the topics generated from the article to filter out the true negatives.

Data Extractor from 2D Plots : Software Development

- Worked in a team of 15 members in Inter-Hall Software Development Competition to build a graph extractor that detects plots in any PDF and digitises the graphs.
- Built a module which digitises the plot lines using the pixel values from the graphical(BW) image by aptly scaling them to the given range of the axes imported from the OCR module using **Python Imaging Library**.

TECHNICAL SKILLS

- **Programming Languages:** C++, Python, C, SQL
- **Libraries and Frameworks:** Scikit-Learn, Graphlab, TensorFlow, Keras, OpenCV, Django
- **Software:** MATLAB, Tableau, Power BI, Eclipse, PyCharm
- **Other Tools and Documentation:** Amazon EC2, Git, LaTeX, UML Diagrammer

COMPETITIONS

- **Inter-IIT Technical Meet: Optimal Bidding | 2018:** Selected as the **captain,** from a pool of 400+ applicants, of the **gold-winning data science team** to represent IIT Kharagpur in the technical tournament attended by **19 IITs.**
- **American Express: Analyse This - Data Analytics Competition | 2017:** Milestone-2 runner up, **secured 2nd position** in the leaderboard **out of 1400+ teams** participated from top-tier colleges in India. Got offered an opportunity to interview with American Express for **full time position.**
- **General Championship - IIT Kharagpur | 2016:** Captained LBS Hall of Residence's **Data Analytics team** to secure **2nd position** in the annual event attended by a total of 20 teams.

POSITIONS OF RESPONSIBILITY

Teaching Assistant | Programming and Data Structures | IIT Kharagpur | Aug '17 - Present

- Tutored the 1st year students on the basics of Programming and Data Structures on a weekly basis at IIT Kharagpur.
- Co-authored the tutorial and assignments sheets for the students on the basic concepts of Programming and Data Structures.

Captain | Data Analytics Team | IIT Kharagpur | Aug '16 - July '17

- Led a team of 15 members in the inter-hall data analytics events carried out throughout the academic year.
- Conducted tutorial sessions on machine learning using python to help members develop deep understanding in the field.

General Secretary | Code Club | IIT Kharagpur | Aug '15 - July '16

- Spearheaded the development of tutorials on various fields of Computer Science to help student community in the institute.
- Conducted regular workshops to familiarise the members with Competitive Coding and Application Development.
- Collaborated with Investment Organisations to organise five coding events which saw significant number of participants.