Abhinay Kumar

+91-8756751942 | abhinaykdn1@gmail.com

Education

Indian Institute of Technology, Kanpur **8.0/10.0**
Bachelor of Technology, Electrical Engineering *2015 – 2019*

Work Experience

Senior Business Associate, Gartner *June 2019 – Present*
Data Science Specialist *Validation & Moderation, Gartner Peer Insights*

- Responsible for scaling efforts of V&M through development of machine learning and automation pipelines
- **Predictive Modeling** : Developed an explainable model using **XGBoost & SHAP** with an accuracy of **90%**
  - Pilot was conducted using shadow & canary deployment frameworks, pilot feedback was included into final results
  - The model led to increased detection of fraudulent cases by **50%** and reduction in moderation time by **15%**
  - The refactored development scripts and documentation were handed-over to engineering team for deployment
- **NLP & Deep Learning** : Developed models for downstream review classification & market relevance tasks
  - Developed a **GRU** architecture using **GloVe** embeddings, training it on text corpora of all previous reviews
  - The model achieved **90%** accuracy for review's market classification task and was integrated into the pipeline
  - Fine-tuned pre-trained **RoBERTa** architecture with a classification head to obtain the market-relevance rankings for reviews belonging to products lying in multiple markets on Peer Insights, achieved top-2 accuracy of **70%**
- **Fraud Detection** : Automated detection & retro-active cleanup of Duplicate Accounts on Peer Insights platform
  - Conducted analysis for all reviewer's data to retro-actively **detect & block 10% duplicate accounts**
  - Automated & integrated the rule-based preventive detection algorithm into the moderation pipeline
  - The algorithm detected the multiple linked accounts and established the intent of user to derive the final decision
- **Process Automation** : Continuous improvements in process pipeline to improve review moderation efficiency
  - Transitioned into the role during deployment of version-1 rule-based checks; collaborated with the moderation & engineering team on **last phase developments & post deployment UAT**
  - Collaborated with moderation team to outline & automate reviewer company information using 3rd party data
  - Developed **reviewer query investigation & resolution automation** Python script in collaboration with Customer Experience Management team; led to decrease in resolution time by **60%**

Business Analytics & Reporting *Advanced Analytics, Gartner Business Services*

- Strategic partner to **Gartner Operating Committee**, driving client retention improvements through analyses
- Led the weekly client outreach discussions & analysis which led to **15% increase in client outreach**
- Developed **Service Enabler Dashboard** using PowerBI to drive standard coaching practices within services
- Designed python script to automate the process for centralized retention & engagement workbook, ensuring uniform analysis & discussions across Sales, Services & Products; led to reduction of manual efforts by **100 hrs/month**
- Designed & automated reachout-to-engagement conversion algorithm in PowerBI to help identify performance bright spots & drive adoption of standard practices; increased the positive client value capture from **5% to 12%**

Techno-Managerial Intern, ITC *May 2018 – July 2018*
Foods division

- Objective : Design & implementation of **Industry 4.0** modules on the finger snacks (Bingo) production line
- Designed and deployed automated **KPIs** for measurement of live **Line Throughput** and **Downtimes**
- Statistically modelled the line throughput at Bingo production line in **kg/hr** using station line variables

Projects

Cassava Leave Disease Classification | *Python, Image Classification, Ensemble* *Dec. 2020 – Jan. 2021*

- Developed a CNN-based image classification model using 21k labeled images with an accuracy of **88.56%**
- Trained an ensemble model using **Xception** and **Efficientnet** architecture to obtain accuracy in excess of 80%
- Mitigated training data noise using a custom **bi-tempered logistic loss** (*https://arxiv.org/abs/1906.03361*)

Indian Premier League | *Python, Data Analysis* *Nov. 2020*

- Analysed 801 IPL matches to identify the best performers in terms of **averages, strike-rate & runs-per-over**
- Compared the best performers using unconventional metrics like **effectiveness** in powerplay and death overs, **batsmen-to-bowler match-ups** & **effect of venue** on a player's performance

Technical Skills

Tools & Languages: *Python* | *C++* | *SQL* | *PySpark* | *PowerBI* | *Excel* | *Git*
Libraries: *Tensorflow* | *HuggingFace* | *Scikit − Learn* | *NumPy* | *Matplotlib* | *C++STL*