

Siddharth Kapoor

Sid.kapoor.che@gmail.com | +1(407) 435 9141 | [LinkedIn](#)

SUMMARY

As an accomplished Senior Data Scientist, I excel in driving data-driven solutions across finance, insurance, and travel industries. With expertise in statistical learning algorithms, anomaly detection, and ML pipeline development, I'm passionate about using data to optimize business processes and decision-making.

WORK EXPERIENCE

Senior Data Scientist, LexisNexis Risk Solution, Atlanta

Mar'19–

Present

- Designed and implemented anomaly transformer, autoencoder, and contrastive learning-based models to detect anomalies and pattern changes in engineering systems performance. **This led to automatic failovers from Production to DR server.**
- Utilized Kernel SHAP-based Shapley values to pinpoint features contributing to anomalies and Implemented Integrated Gradient, a method for feature attribution in deep neural networks using TensorFlow framework. **The Shap values provided insights for slowness in the Production environment.**
- Developed an ensemble model using PCA, Kmeans, DBscan and isolation forest to identify peak or extreme latency in product queries performance within Roxie systems in the DR environment. **This project reduced several man hours for analysis.**
- Developed automatic machine learning training pipelines which takes the input from the forms and schedule the training of the models for the ensemble model for DR project.
- Developed an Airflow pipeline encompassing Data ETL processes to collect and aggregate data from diverse systems such as ELK, SQL Server, and Prometheus, as well as Machine Learning, feature attribution, and reporting pipelines for all the projects.

Research Assistant, Georgia State University, Atlanta

2017–

18

- FINRA Dispute Prediction: Applied Word2Vec for similar text search and n-grams for feature building purposes to predict claimant awards, using SVM and Decision Tree classifiers.
- Court Misclassification (DOL): Predicted judge opinions in employee-employer disputes using logistic regression, Word2Vec, NLTK, and Spacy to analyze 500k documents, identifying decision patterns and key lawsuit reasons.

Payment and Product Analyst Intern, PayPal, Bay Area

May–Aug'18

- Billing Agreement Product: Performed funnel analysis on the Card Updater Business process to evaluate its effectiveness, pinpoint customer segments with high transaction declines, conducted event analysis, and developed a Tableau dashboard for visualization

Team Lead- Analytics, Emirates NBD, Dubai, U.A.E

2015–

17

- Devised a customer profiling and segmentation approach using MySQL to identify at-risk segments by analyzing customer behavior, payment patterns, bank balance trends, salary trends, and expenditure increment trends. **This helped in assigning proper customers to service specialist.**
- Designed a SQL-based method for calculating customer contactability and improved the numbers by incorporating with inbound call data. **The contactability increased which led to better account recovery.**

Product Analyst, OptionTown, Delhi, India

2013-14

- Developed an algorithm that analyze the customer travel data and compress it into a string and correlates with the purchase trends
- Created a text mining algorithm to assess consumer behavior from the history trend and implement a smart emailing strategy aimed at reducing attrition

EDUCATION

- Master's in Computer Science, **Georgia Institute of Technology** July'21-Present
- Master's in Analytics, **J. Mack Robinson College of Business** July'17-Dec'18
- Bachelor's in Engineering, **Indian Institute of Technology, Varanasi** July'09-Jun'13

TECH STACK

- Skills:** Machine Learning, Deep Learning, Data Analysis, Predictive Modelling, and Statistics
- Languages:** SQL, Python, VBA, R
- Tools:** MySQL, Jupyter Notebook, Power BI, Tableau

ACCOMPLISHMENTS

- Extensively employed advanced techniques such as Anomaly Transformers, Autoencoders, Contrastive Loss and Machine Learning ensemble methods to identify anomalies in engineering systems, contributing to improved performance and system reliability
- Utilized Kernel SHAP-based Shapley values and Integrated Gradient methods to pinpoint the most critical features contributing to anomalies, enabling better understanding and interpretation of complex models

