

Soumyadeep Ray

Stillwater, OK 74075 | soumyadeep.ray@okstate.edu | (405)- 385-4172 | [in/Soumyadeep-Ray](https://www.linkedin.com/in/Soumyadeep-Ray)

EDUCATION

Master of Science in Management Information Systems -Data Science

Expected Dec,2022

Oklahoma State University, Stillwater, OK

GPA: 3.75

- **Coursework** : Data Science Programming, Statistics for Data Science, Data warehousing, Predictive Analytics

Bachelor of Technology in Civil Engineering

June 2017

Maulana Abul Kalam Azad University of Technology, India

GPA: 3.75

- **Coursework** : Data Structures, Linear Algebra, Calculus, Statistical tests

TECHNICAL SKILLS

Programming Skills:	Python(Numpy, Pandas, Scipy, Matplotlib, MLib, plotly, scikit-learn, Keras, Tensorflow, Pytorch), R, Java, SAS, C++
Tools and Frameworks:	AWS(S3, EC2), Big Query, Power BI, Terraform, Jenkins (CI/CD), Git, Airflow, Luigi, Hadoop
Data Science Skills:	Supervised Learning (Regression, Classification, Decision Trees), Unsupervised Learning (Clustering, PCA), Ensemble Modeling, NLP, Reinforcement Learning, Hypothesis Testing, A/B Testing, Topic Modeling, Recommendation Systems, Neural Networks, Data Visualization

EXPERIENCE

Data Science Intern – Marriott International

Jun 2022 – Aug 2022

- Built training data pipelines using Spark and Snowflake that consisted of data of 32 million customers having 1000+ features (profile, reservation, digital click, and loyalty information) and stored as parquet files in AWS S3.
 - Developed Propensity Models to identify the customers that have higher likelihood to book reservations.
 - Optimized ML models using Gurobi Solver thereby increasing the overall performance of marketing campaigns by 2%.
- Tech used – PySpark, Python, Snowflake, Airflow, AWS (EC2, S3), Git, Terraform, A/B Testing

Graduate Research Assistant | Oklahoma State University

Jan 2022 - May 2022

- Extracted and processed data using CDC Wonder API to build a comprehensive MySQL database resulting in 95% accurate result on testing period.
- Performed statistical analysis to determine the wheat producing counties in Oklahoma based on the Z scores assigned by comparing it with national average wheat production.

Software Engineer |Tata Consultancy Services Limited

Dec 2017 – Jun 2021

- Managed an Agile development team of 3 associates in terms of processes, functional and data requirements.
- Redesigned the operational process to identify fraudulent claims, enabling a 25% increase in target resolution, using clustering and tree-based models.
- Created Tableau Dashboards to present information to stakeholders and visually monitor KPIs and insights.
- Practiced end-to-end application development with all stages of Software Development Life Cycle (SDLC): design, analysis, coding, and testing.
- Utilized JavaScript to produce high performance, user-friendly web interface reducing user clicks by 20% during navigation.
- Automated the data import from excel to MySQL database using Python scripts reducing the manual efforts by 30%.
- Created and tuned the ETL workflows in Informatica Power Center that reduced the run time of full data loads from source to target systems by 5 hours.

ACADEMIC PROJECTS

Predicting Credit Card Approvals

Apr 2021 – Jun 2021

- Utilized Python to tackle the missing value imputations and preprocessed the dataset by using one hot Encoding, label Encoding and performed featured scaling.
- Built a machine learning model for numerical data to predict an individual's credit card application expectance using NumPy, pandas, Scikit Learn, Seaborn and yielded a classification accuracy of 84%.

Sentiment Analysis of Soccer Teams to understand team performance and mentality

Sep 2021

- Scrapped Twitter to collect 1.6 million tweets regarding the soccer teams and their players and assigned sentiment scores using Vader Sentiment Analyzer. Implemented bag-of-words, stemming and lemmatization.
- Experimented with different classification models such as Naive bayes, SVC, Random Forest to classify our sentiments as positive or negative; best model – Naive bayes with accuracy 0.80.

HONORS

Spot on Award | Tata Consultancy Services Limited

Aug 2018, Feb 2020, Mar 2021

- 3-time winner for excellent performance and contribution to the team.