Sourav Mondal, Ph. D

Researcher, Data Science practitioner

SUMMARY

I am a Physics researcher experienced with complex computational algorithms and statistical analysis of scientific data. My field of research has been Lattice Gauge Theory which emphasizes the close connection between Quantum Field Theory and Statistical Mechanics. Currently, I am trying to use my expertise to solve business problems using machine learning and Deep Learning algorithms. Looking forward to getting an appropriate role to use my experiences and skills.

9867962993

sourav.mon.sm@gmail.com

www.linkedin.com/in/dr-sourav-mondal/

https://github.com/souravmondal-git

Kolkata, West Bengal

KEY SKILLS

Programming Languages : Python, SQL, C

Python packages: Scikit-Learn, SciPy, NumPy, Pandas, Matplotlib, Seaborn, Tensorflow , NLTK.

Machine Learning Algorithms : Linear & Logistic Regression, Decision Tree, SVM, KNN, Random Forest, XGB, K-means clustering, PCA, Isolation Forest .

Deep Learning Algorithms : ANN, CNN, Auto-Encoders, RNN

Other Quantitative Skills: Data Science, NLP, Statistical Analysis, Physics, Monte Carlo Simulations, Mathematical modeling.

EDUCATION

2016	**Ph. D. (Theoretical Physics)** *IACS, Kolkata* (University of Calcutta)
2008	**M. Sc (Applied Physics)** *IIEST, Shibpur*
2006	**B. Sc (Physics)** *University of Calcutta*
2003	**Higher Secondary** *W.B.C.H.S.E*

WORK EXPERIENCE

2021 - 2018	**Postdoctoral Research Associate** **IISc, Bengaluru**

Used statistical analysis and mathematical modeling of data generated by Monte Carlo simulations to estimate QCD thermodynamic variables at finite temperatures and chemical potentials using an improved method. Collaborated with HotQCD collaboration and used the computational facility of Bielefeld University, Germany.

2018 - 2016	**Postdoctoral Research Fellow** **TIFR, Mumbai**

Used statistical analysis of data generated by Monte Carlo simulations to estimate Physical observables like the mass of a singly heavy baryon which was discovered by the LHCb experiment in 2017.

2016 - 2009	**Research Fellow(*Ph.D program*)** **IACS, Kolkata**

Generated SU(3) gauge field configurations using Markov Chain Monte Carlo algorithms on four dimensional hypercubic lattice. Using those configurations calculated Glueball (particle yet to be discovered experimentally) mass. Developed algorithm to reduce noise in Glueball mass spectrum.

KEY DATA SCIENCE PROJECTS

Movie Recommender System: Used Movie-Lens Data, consisting of user IDs, movie IDs, ratings given by users to different movies, to build a Recommender System. "Surprise" library has been used to split the data into train and test sets. Collaborative filtering, CoClustering, and Matrix factorization based algorithms have been used. Matrix factorization based algorithm SVD++ produced minimum RMSE. Final model has been built using SVD++ to suggest the top five unseen movies to each user.

Predicting Employee Attrition: Employee data of joining date, educational qualification, last working date, total business value, quarterly ratings, etc were given in a Hackathon. Problem was to predict whether employees will leave the company within the next two quarters. Target feature is created using the data whether the last date of working is available or not. Used Random Forest algorithm to build final model with accuracy score 0.75. Leader board score for that model was 0.6539.

Credit Card Fraud Detection: Credit card transaction data consisting of 29 independent feature variables used for this project. The problem was to find whether any transaction characterized by all these features is fraudulent or clean. Although the data set had a target variable, I did not use that to build the model. Used Auto-Encoders to build an unsupervised anomaly detection model. My model correctly detected 86.8% of the fraudulent transactions.

Airlines User Reviews Sentiment Analysis: Used sentiment analysis to predict sentiments of user reviews of an airlines company. Labeled data for the sentiments were classified among three classes - positive, neutral, and negative. Text processing was done using NLTK and regex. Word embedding was done by Word2Vec. LSTM model predicted the sentiments with 77% accuracy.

E-mail Spam filtering: Given the dataset of 5171 e-mails, I have used NLTK and Regex libraries to preprocess the email text. Used Word2Vec model for word embeddings. And finally, build the model using KNN classifier with 94.3% accuracy.

CERTIFICATION

2021	CERTIFIED MACHINE LEARNING WITH PYTHON EXPERT by Decodr Technologies Pvt. Ltd.