DEBODEEP BANERJEE

Viale della primavera, 15, Roma- 00172, Roma, RM, Italy
Contact:+39-3510511586 **E-mail: debodeepbanerjee33@gmail.com**
Github: Debodeep94

Objective
I want to explore new dimensions of the world of statistical inference and data science. Through the ages, especially in the regime of Statistical data analysis, the inferential methods have played a revolutionary role. On that note, my interest lies in the field of statistical inference, data mining and machine learning. My aim is to work on this field in a more rigorous way so that my contribution enriches the subject as well as its application in real-life cases. Apart from that, playing with various visualization techniques is another point of interest of mine.

Areas of interest

•Bayesian inference
•Data visualization
•Natural language processing
•Directional data analysis

Education

•Laurea Magistrale in Data Science (Masters in Data Science)
Sapienza University of Rome (2019-)
•Master of Science in Statistics
Department of Statistics, University of Madras (2016-2018)
•Bachelor of Science, Statistics (Honors)
Asutosh College, University of Calcutta (2012-2016)
•Higher Secondary Education
Uttarpara Government High School, West Bengal Council of Higher Secondary Education (2010-2012)
•Secondary Education
Baranagore Ramakrisna Mission Ashrama High School, West Bengal Board Secondary Educaion (2010)

M.Sc.(in Statistics) Research project

Title: A study on classical inference on single binomial proportion
Duration: October, 2017- May, 2018
Supervisor: DR. Subbiah M. (Former Associate Professor, Presidency College, Chennai, India)
Short description:
We revisited the problem of interval estimation of a binomial distribution.The focus has been fixed on the justification of choice of several confidence intervals based on their respective coverage probabilities. The performances of the confidence intervals have been analyzed with help of coverage probabilities and expected run length. Further, proposition of new interval has been placed and it has been compared with the help of coverage probability and expected run length.

Technical report

Title: Comparative Analysis of Two $1 \times j$ contingency table under Bayesian perspective
Authors: Deboddep Banerjee, Trina Sahoo, DR. Subbiah M. and DR. M.R. Srinivasan
Duration: March, 2017- February, 2018
Publication: University department library, Department of Statistics, University of Madras, February, 2018
Short description:
This study has exploited these aspects into categorical data of two one-dimensional variables each with same J levels of categories. The notion focussing on the ratio of independent Beta distribution is considered using closed form approach with Gauss hypergeometric function and Monte-Carlo techniques. The entire approach is illustrated with a primary data set that aims to study the impact of gender on perceived important social issues; essential computations are carried out using R program. The modelling advantages of Bayesian approach has been studied and the results are directly interpreted regarding the context of the problem.

Paper presentation

•Presented Paper in International Conference on Theory and Application of Statistics and Information Science, 2018
Topic- Comparative Analysis of Two 1×j contingency table under Bayesian perspective
•Presented paper in National Conference on Dimensions of Sustainable Economic and Business Development organised by Marian College, Kuttikkanam, 2018
Topic- CSR and Sustainable Development: An Analysis of Practices in India

Academic project(Click on the title for Github link)

1 Websraping, LDA and forecasting

In this project I scraped the data from a website where all the speeches of several central banks' board members are stored in textual formats. I scrapped their speeches based on USA and China. Some concepts of NLP was used for preliminary data analysis. Then I implemented Latent Dirichlet Analysis and finally I addressed a forecasting problem with LSTM and LSTM autoencoder.

2 Stock market prediction using twitter data

In this project, I encounter a stock market prediction problem with the data of 2015 to 2021 of ONGC (Oil and Natural Gas Corporation). In this project we have implemented LSTM model in order to predict the future data. Further, specifically in the stock data, the volatility of stock data highly depends on news and/ or social media. Combining the scores obtained from the sentiment analysis of the tweets and the stock data, another LSTM model has been developed.

3 Model interpretability using LIME

In this project I considered a classification task using CNN and implemented LIME to address the concept of model interpretability. I used a dataset of archaeological buildings with 10 classes.

4 Fire Detection using convolutional neural network

We collected the data from Kaggle and CVPR lab. The aim was to classify between images with fire and images without fire. For this task we compared between three models, viz. CNN, Bayesian CNN and Resnet50.

5 Performance of Knowledge Organisation and Human Action using Probability Graphical Model and Efficiency Analysis

We did a Literature review of the paper obtained from Scopus and WOS. Then we introduced an application of PRISMA model and Ontological Modelling using *Eddy* and *Protege*. We made the visualization using Biblioshiny, VOSviewer.And finally we applied Probability Graphical Model and Efficiency Analysis.

6 Linear regression using JAGS

In this project we compared the performance of two statistical models. The first one is multiple linear regression model and the second one is bayesian model using JAGS.

7 Covid situation visualization

It is a self interested project. I tried to create some visualization(both static and animated) which depict the contemporary situation of India in the covid situation.

Internship

Institute: Indian Statistical Insitute
Topic: Data Analysis using Bootstrapping and Logistic Regression.(May, 2017- June, 2017)

Certificate program

Program title: Certificate program on Machine learning and Artificial intelligence
Institute: Madras School of Economics
Duration: June, 2018- August, 2018

Program title: Certificate course on Python
Institute:FITA academy
Duration: December, 2018- January, 2019(tentatively)

Work experience

Organization: Baker Hughes
Designation: Artificial Intelligence intern
Title: Deep learning on data acquired by turbomachinery
Duration: July, 2021-Jan, 2022

Organization: SPI Global Private Limited
Designation: Subject Matter Expert, Statistics
Duration: May, 2018-August, 2018

Software proficiency

•**Languages:** Statistical applications of C language, Python, R
•**Statistical analysis software:** Minitab, SPSS
•**MS Office suit:** MS Office (MS Word, MS Excel, MS PowerPoint, MS Access)
•**Other(s):** LaTeX, SQL, Power BI(beginner)

Awards and achievments

•Secured first place in the Inter-Departmental Quiz Competition, organized by the Department of Library and Information Science, University of Madras.
•Secured first place in quiz in the Inter-Collegiate competition, organized by the Department of Statistics, Madras Christian College.

Key skills

•Language proficiency in Bengali, Hindi, English(TOEFL Score 98).
•Ability to meet deadlines.
•Eager to learn new concepts.

Hobbies

•Reading books (Mainly non-fiction)
•Writing stories, articles, poems