DR. TASHA GAUTAM
DATA SCIENTIST
16.TASHA@GMAIL.COM | +1 434 760 9111 | TORONTO, CANADA
LINKEDIN, PORTFOLIO, GITHUB, GOOGLE SCHOLAR

Summary

- Ph.D. in physics with over **5** years of hands-on experience in data science projects, showcasing a strong foundation in quantitative research and problem-solving.
- Skilled in machine learning, data analysis, visualization, statistics, and adept at deriving insights from complex, unstructured data, as evidenced by **4** lead-author scientific publications demonstrating data science expertise, independence, and innovation.

Technical skills

Programming Languages	Python (6 yrs) | Bash (6 yrs) | C++ (5 yrs) | SQL (2 yrs)
Data analytics	Pandas | Numpy | Scipy | Seaborn | Matplotlib | Jupyter | Git | Time-series analysis
ETL pipelines (Python | Bash)	Slurm workload manager on HPC cluster | Container creation: Docker, Singularity | Cloud computing
Machine Learning	Scikit-learn | Regression | SVM | Random Forest | Naive Bayes | K-means | Natural Language Processing (NLP) | ARIMA, SARIMA models

Key Experience

Natural Language Processing (NLP) - Predictive and Sentiment Analysis

Portfolio project

- Performed predictive analysis using **4** Machine Learning classification models on customer reviews of products by a clothing brand. Used word embedding like Bag-of-words model to achieve **>86%** accuracy.
- Executed sentiment analysis on **3** months of ChatGPT tweets using Part-of-Speech tagging and NLTK's TextBlob and Vader sentiment analysis. Classified positive and negative sentiments and got an accuracy of **94%** on the prediction.

Time-Series Forecasting

Portfolio project

- Performed time-series analysis on **7** years of Stock Price of FANG companies using ARIMA and SARIMA models. Forecasted future stock price for the next **5** months with **90%** C.I.

Data analysis and modeling

NRAO, 01/2023 - Present

- Integrated new feature to 'timing analysis' GitLab repository of our collaboration to develop new models for **68** star systems, performed chi-square, F-test, and z-test to eliminate outliers and identified important features in the model.

Statistical and Predictive Analysis [Published paper link]

MPIfR, 08/2021 - 12/2022

- Analysed **10 years of data** (**> 10 TB**) by implementing statistical techniques: chi-square minimization, regression, and likelihood estimation, detected **3** new phenomena in a star system, achieved **5-sigma** constraint on star's mass.
- Conducted predictive analysis with data simulations using griding and contour techniques and found **up to a factor of 6** improvement in the parameters.

ETL pipeline | Time Series Analysis [Published papers links: 1, 2]

MPIfR, 08/2018 - 12/2021

- Built an efficient python based pipeline (**24** times faster) for High Performance Computing (HPC) Cluster by implementing parallelization in multiple steps, as a result **discovered 6** new neutron stars.
- Created containerized environments using docker and singularity containers, and implemented data analysis on **100+** time-series datasets including cleaning, analyzing, visualization, down-sampling, and faint periodic signal searches.
- Reduced computing time by deploying jobs on cloud computers using SLURM batch system, coordinated parallel computation across **>80** nodes with optimized CPU/GPU allocation for efficient data processing with bash scripts.
- Executed statistical algorithms and time-series analysis to derive scientific insights from **>10 TB** of data collected in **2** large surveys and modeled properties of newly discovered stars.

Education

Ph.D. - Astrophysics | 2018-2022
Max Planck Institute For Radio Astronomy (MPIfR), Germany
12 published papers, 350+ citations

Masters - Physics | 2015-2018
Indian Institute of Science Education and Research, India
Awarded certificate of academic excellence

Bachelors - Physics | 2012-2015
University of Delhi, India

Leadership and Management

- **Leadership:** Led **4** data science related research projects | Published **3** first author and **2** second author papers in distinguished international journals | Mentored interns | Representative at International Max Planck Research School for Astronomy and Astrophysics.
- **Teamwork:** Collaborated on **12+** research projects with teams of up to **50** scientists | Proud member of NANOGrav collaboration (**100+** scientists) that facilitated the ground-breaking discovery of gravitational wave background using pulsars.
- **Communication : 3-4** public presentations per year since 2018 to technical and non-technical audience of up to **30-40** people | Teaching Assistant in a Master's level course at University of Bonn.
- **Organisation:** Organised weekly seminars for doctoral candidates at International Max Planck Research School.