TASHA GAUTAM, Ph.D.
Data Scientist

Toronto, Canada • 16.tasha@gmail.com • +1 (437) 446 9172

LinkedIn • Portfolio • Github • Google Scholar

SUMMARY

- With a Ph.D. in Physics and **5** years of experience in data science along with a strong foundation in quantitative research, I bring a unique perspective and a creative acumen to address real-world challenges.
- Skilled across the entire data cycle from cleaning to analyzing to deriving insights using statistics, I also bring project management expertise and communication skills, as evidenced by **12**+ publications and **20**+ talks.

TECHNICAL SKILLS

Programming Languages	Python (6 yrs) • Bash (6 yrs) • C++ (5 yrs) • SQL (2 yrs) • Git (5 yrs)
Data analytics	Pandas • Numpy • Scipy • Seaborn • Jupyter • Matplotlib • Tableau • PowerBI
ETL pipeline	Slurm workload on HPC cluster • Docker, Singularity • Cloud computing • GCP
Machine Learning	Scikit-learn • TensorFlow • Regression • Classification • Clustering • SVM • Random Forest • K-means • NLP • Time Series Analysis: ARIMA, SARIMA, Prophet

PROFESSIONAL EXPERIENCE

NATIONAL RADIO ASTRONOMY OBSERVATORY   **Charlottesville, Virginia, USA**
Research Associate, Data Science Expertise   **JAN - NOV 2023**

- Developed an automated python pipeline to create new models for **68** stars, performed statistical tests to eliminate outliers and identified new features from the data, got >**150%** improvement in the measurements.
- Led a team of **8** to analyze **TBs** of data acquired over **13** yrs, implemented time-series analysis using regression and Bayesian techniques, as a result discovered a novel phenomena in a star and published a lead-author paper.

MAX PLANCK INSTITUTE FOR RADIO ASTRONOMY   **Bonn, NRW, Germany**
Research Assistant, Data Science Expertise   **2018-2022**

- Analyzed **10 years of data** (> **10 TB**) using statistical techniques, detected **3** new phenomena in a star system, and achieved a **400%** improvement in the measure of mass for a star.
- Accomplished **6-fold** improvement in the parameters from predictive analysis of the data and published results in a lead-author paper.
- Created an efficient ETL python pipeline and a containerized environment (docker and singularity), analyzed **100**+ time-series datasets to **discover 6** new neutron stars.
- Coordinated parallel computation across >**80** nodes with optimized CPU/GPU allocation on cloud computers using SLURM, reduced computation time by **24** times with enhanced performance of the pipeline.
- Executed statistical algorithms and time-series analysis of >**10 TB** of data collected in **2** large star surveys to model properties of newly discovered stars, published results in two lead-author papers: 1,2.
- Led teams of **10**+ scientists and published results in **4** papers, executed within a timeline of **6** months.

SELF-GUIDED PROJECTS   **Toronto, ON, Canada**
Predictive Analysis of customer reviews for an e-commerce clothing brand | **NLP**

- Performed predictive analysis, trained Machine Learning classification models on customer reviews of products from a clothing brand. Used word embedding methods to achieve >**86%** accuracy.

Sentiment Analysis of ChatGPT tweets over three months | **NLP**

- Executed sentiment analysis on **3** months of ChatGPT tweets using Part-of-Speech tagging, NLTK's TextBlob and Vader tools. Classified positive and negative sentiments, trained ML models and got an accuracy of **94%**.

Analysis and Forecasting of Stock prices of FAANG companies | **Time Series Analysis**

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

EDUCATION

UNIVERSITY OF BONN   **Bonn, Germany**
Ph.D. in Astrophysics   **2018-2022**
12 published papers, **350**+ citations, Magna Cum Laude

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH   **Mohali, India**
Master of Science in Physics   **2015-2018**
Awarded certificate of academic excellence

LEADERSHIP AND MANAGEMENT EXPERIENCE

- **Leadership:** Led **4** data science research projects with cross-functional teams of **10** • Published **4 lead author** papers in distinguished international journals • Led **8** telescope proposals (reports) • Mentored **2** interns • Representative at International Max Planck Research School (IMPRS), Germany.
- **Teamwork:** Collaborated on **12**+ data science projects with teams of up to **50** scientists • Active member of NANOGrav collaboration (**100**+ scientists) that facilitated the ground-breaking discovery of gravitational wave background.
- **Communication:** Delivered **20**+ public talks to technical and non-technical audience of up to **40** people • Teaching Assistant in Master's course at University of Bonn, Germany.
- **Organization:** Organised and chaired weekly seminars and yearly conference at IMPRS, Germany for researchers.