

## Summary

Senior Data Scientist at Red Hat. Total 6+ years of experience in Data Science. Excellent knowledge of Data Science and Data Analytics. Proficient in Machine Learning, Optimization, Mathematics, Statistics, Forecasting, Natural Language Processing, Deep Learning, LLM & Generative AI, etc.

## Work Experience

- Oct 2024–Present **Senior Data Scientist, Red Hat**
- Nov 2023–Oct 2024 **Senior Data Scientist, Fractal Analytics**
  1. Leading Data Analytics Team for a CPG Client. Thought Leadership, Creating models for price & weight elasticity, Demand Transference, and Innovation for Price Pack Architecture. Scaling this project to different markets. Creating features for different markets for modelling. Create a sprint plan for the Agile environment. Implement Data Science enhancements & Data refinement in the existing framework. Implemented MLOPs framework, and parallel processing to improve code runtime.
  2. Forecast Reconciliation for sales volume of Beer for a CPG client.
  3. Worked on Data Filtering, Sentiment Analysis, Topic Modelling, Identifying Pain & Passion points for oral care data using NLP & GenAI,
- Aug 2019–Nov 2023 **Data Scientist, Red Hat**
  1. Developed a Bayesian model for services consumption probability to improve seller knowledge for lead generation and portfolio growth. Backtesting the model for four quarters. Implemented GitLab CI/CD pipelines for Python on OpenShift, enabling automated processes. Created data for the recommender system to establish service-offering rules.
  2. Achieved accurate forecasts (MAPE) for Commissions(7%), Sales(3%), Expenses(6%), and Billings(4%) globally using traditional ML and ensemble methods. Utilized Hierarchical and Grouped time series with Forecast Reconciliation. Visualization in Tableau dashboard for comparison with earlier methods and actuals. Employed Champion-Challenger framework for model comparison. Created external regressors for multivariate modelling from business understanding.
  3. Successfully forecasted Par-rate within 15 days, achieving 1.15% MAPE globally for various hierarchies.
  4. Developed Python library for Time Series Forecasting.
  5. Conducted Finance analytics. Automated cash flow data collection through SQL.
  6. Worked on highly imbalanced data (97:3 ratio), implementing techniques such as EDA, Data Visualization, Feature Engineering, ML algorithms, model validation, backtesting.
  7. Managed quarter-end activities and POCs with stakeholders.
  8. Ensured data validation with the data engineering and reporting team.

- Dec 2018 - **Data Science Intern, CRISIL**
- Aug 2019 Developed Credit Risk Modelling Tool(GUI) for scorecard generation, integrating Python Dash with Machine Learning, Statistical, Financial & Optimization techniques. Employed model-agnostic techniques to explain black-box ML algorithms (**M.Tech thesis**).
- July 2015 - **Graduate Apprentice Trainee, SKF India Ltd**
- July 2016 Led team as supervisor/channel team engineer.

## Academic Qualifications

- 2017–2019 **Centre for Modeling and Simulation, M.Tech in Mathematical Modeling and Simulation, Savitribai Phule Pune University**
- **CGPA - 8.98/10**  
Specialized in Machine Learning. Proficient in Mathematical & stochastic Optimization, Differential Equations, Linear Algebra, Numerical Methods, Calculus, Statistics, Operational Research, NLP, Deep Learning, etc. Expertise in implementing Machine Learning, Optimization algorithms, and Numerical Methods using Python & R.
- 2011–2015 **Pimpri Chinchwad College of Engineering, Bachelor of Engineering, Mechanical, Savitribai Phule Pune University**
- **BE Aggregate: First Class (63.86 %)**  
Majors in Applied Mathematics, Machine Design, Dynamics of Machinery, Thermodynamics, Heat Transfer, IC Engine, Mechanics & Fluid Mechanics, Theory of Machines, etc.

## Ongoing Research Work

- Machine Learning classifier Ensembles using Ant Colony Optimization & Black Hole Algorithm
- New proposed Feature Selection using Black Hole & Genetic Algorithm
- Minority class oversampling with Machine learning methods.

## Academic Research

- **Feature Selection for Multi-label dataset**  
Reducing multi-label dataset dimensionality through feature selection. Utilizing filter methods like Multi-label Informed Feature Selection, Robust Feature Selection, and Mutual Information.
- **Detection of specific antibodies/allergens using Machine Learning**  
**Collaboration with Bioinformatics Centre, Pune University.:** From AllerBase, a comprehensive knowledge base of allergens. Classification of allergen-related features such as IgG, IgM, IgA, and IgE using Machine Learning.
- **Ayurveda Prakriti Classification using Machine Learning**  
**Collaboration with C-DAC, Pune University Campus:** Applying Machine Learning for Ayurveda Prakriti (Dosha) Classification. Addressing it as a Multi-class and Multi-label problem.
- **Forecast Indian summer monsoon rainfall using ensemble techniques**  
**Collaboration with Indian Institute of Tropical Meteorology (IITM), Pune:** Using ensemble techniques of Machine Learning for prediction of Indian summer monsoon rainfall.
- **Classification of atomic clusters using Machine learning**  
**Collaboration with CSIR-National Chemical Laboratory, Pune:** Machine Learning-based classification of atomic clusters of Gallium considering shape, atomic distances, and geometrical properties. Utilizing both supervised and unsupervised Machine Learning methods.

## Data Science Competitions

- **Loan Prediction Competition by Analytics Vidhya**

Given imbalanced dataset for prediction of whether a person will get a loan or not for which they have applied. Used ML models with EDA, feature selection & engineering, Data Imbalance techniques, Grid search, etc. **Rank: 15/81895      Accuracy : 82.638%**

- **WNS Analytics Wizard 2018**

Predict whether a potential promotee at checkpoint will be promoted or not after the evaluation process. Used ML models with EDA, feature selection, Data Imbalance techniques (Imbalance ratio: 92:8), Grid search, etc. **Rank: 638/3937      F1 Score : 0.488**

- **StumbleUpon Evergreen Classification Challenge**

Built a classifier which will evaluate a large set of URLs and label them as either evergreen or ephemeral. Used EDA, Text Data Preprocessing, Word Embedding, Feature engineering, Machine Learning models such as CatBoost and Logistic regression, Deep Learning models such as LSTM and BERT

- **Toxic Comment Classification Challenge**

Built a multi-label model which is capable of accurately detecting different types of toxicity like threats, obscenity, insults, and identity-based hate from comments made on social media posts. Used Data Transformation strategies such as Binary relevance, Label PowerSet, Classifier Chains and RaKel to solve multi-label problem. Used EDA, Text Data Preprocessing, Word Embedding, and Machine Learning models such as XGBoost and Logistic regression. Deep Learning models such as LSTM, and BERT.

## Technical Skills

- **Programming Languages:** C/C++, R, Python, SQL, Latex.

- **Python Libraries:** Numpy, Pandas, Scikit-learn, Sktime, Statsmodels, Seaborn, Matplotlib, Keras, Tensorflow, PyTorch, transformers, Spacy, NLTK, regex, Plotly, etc.

- **Certifications:** Specialization in Deep Learning (Coursera), Tableau, Git, GenAI with LLM, etc.

- **Technical Knowledge:** CI/CD Pipeline, Docker, OpenShift, Excel, and GSheet.

## Achievements and extra-curricular activity

- Conducted Scientific Data Mining and Visualization course with hands-on Python as a Teaching Associate with Prof. Jayaraman at Bioinformatics Department, Pune University, from Aug 2024 to Nov 2024
- Conducted Data Analytics course with hands-on R as a Teaching Associate with Prof. Jayaraman at Flame University, from July 2019 to Nov 2019
- Conducted Stochastic Optimization course with hands-on Python as a Teaching Associate with Prof. Kavita Gangal, Centre for Modeling and Simulation, Pune University, from July 2023 to Nov 2023
- Conducting lectures on Python & R Hands-on, Machine Learning and Stochastic Optimization at Centre for Modeling and Simulation, Pune University for 8 Semester (7-8 Lectures/Semester)
- Presented a WEBINAR on "Recent Trends in Machine Learning" at the Department of Electronics and Telecommunication, SKN Sinhgad College of Engineering, Solapur.
- Presented "Feature Selection Using Evolutionary Algorithms" at Computer Science Symposium at Flame University, Pune
- Achieved 3rd position in the Project Competition during MECHFEST-2015 at PCCOE.
- Received the Consolation Prize in the poster competition for the Most Marketable Design at PCCOE.