ANURAG BHATT

DATA SCIENTIST, TSCNET Services GmbH, Munich, Germany
PHONE : +49-173 5626609 • E-MAIL : anuraag.bhatt@gmail.com • Homepage : https://anurgbht.github.io/home

To be out there, in the midst of it, endowed with a capacity for leadership, tenacity and enterprise, and a sense of responsibility, delivering the ends and enjoying what I do.

About Me

5+ years of experience in the field of data analytics. Starting from a theoretical statistical thesis in my Master Program to development, execution, validation and deployment of ML models in power, manufacturing, finance and retail industry.

Education

Integrated B. Tech - M. Tech in Mechanical Engineering and Statistical Methods – IIT Kanpur – 2016
CGPA: 8.2 / 10

Master of Arts in Philosophy
IGNOU – 2018 – 77.2%

Higher Secondary – Central Academy, Dadabari, Kota (CBSE) – 2011
Percentage: 90.6%

Secondary – Kendriya Vidyalaya, Junagadh (CBSE) – 2009
Percentage: 93.6%

Skills

Machine Learning, Modelling, NLP

TensorFlow, H2O.ai, PyTorch

Client Management, Communication

Python, R, MatLab, SQL

C#, C++, Java, HTML/CSS/JavaScript

Presentation/Documentation

Big Data, Hadoop, Hive

Area of Work and Responsibilities

- Development and automation of reporting framework including data storage, analyses and visualization.
- Exploring and leveraging advanced machine learning techniques – like ANN, SVM, GBM, Random Forests, Deep Learning, etc.
- Get involved in direct client engagement when required by client facing roles.

Work Experience

TSCNET Service GmbH, Munich, Germany
Data Scientist *October 2019 – Present*

TSCNET Services turns data into information for a secure, prosperous and sustainable Pan-European synchronized power transmission grid.

Key projects:

- ***Early Warning System*:** Prediction of overloaded transmission grid elements can significantly increase operational security for Transmission Service Operators (TSOs). We extracted relevant features from the available historical generation, consumption, grid topology and weather data to train nuanced models to predict the overloaded elements. The early warning was provided to the TSOs for mitigative action. Insights generated during the project were leveraged to further enhance data quality and gain operational understanding.

- ***Automated Data Quality Assurance*:** Quality of the available data determines the success or failure of machine learning endeavors to come. Under Data Quality Assurance (DQA), we automated report generation and consistency checks on all the data streams and pipelines to ensure robust and reliable data. We integrated all the data streams to the in-house PI Database and developed custom solutions for automated monitoring using PI-SDK. This involved continuous monitoring and outlier detection in the data streams, dashboards and web apps for visualization, and email notifications.

Tata Insights & Quants (Tata iQ), Bangalore, India
Data Scientist *March 2018 – August 2019*
Jr. Data Scientist *August 2016 – March 2018*

Tata Insights and Quants offers multi-sectoral advanced analytics and data engineering solutions using sophisticated predictive analytics and machine learning algorithms.

Key projects:

- **Unsupervised anomaly detection coupled with supervised classification:** Failure tagging in IoT sensor data is frequently absent or incorrect. We used reconstruction error based unsupervised anomaly detectors for accurate tagging of the time series sensor data. This tagging was further used to generate upstream features that were fed to a supervised classifier.

Academic/Personal Projects

- A personal tool for language learning using language translation APIs and NLP, NLTK, in Python.
- A custom decision tree tool that supports editing the nodes and splits, in Python.
- A python tool to convert python model objects into machine implementable C++/C# codes. The class can handle GBM, RandomForest, Keras MLP, XgBoost and SVM models. This becomes essential when deploying the code in production environments.
- AutoIt based python script to automate the extraction of the machine generated data
- Genetic Algorithm: Combinatorial Optimization of thickness of a composite plate using GA
- A GUI based platform for Digital Image Correlation of plates subjected to strain in MatLab
- Imputation of sparsely filled matrices

Academic Publications

- Submitted to archives: A Two-Step Item Response Theory Model for India's Graduate Aptitude Test in Engineering

Academic Conferences

- ICASP-5 & CSSCR-5: The use of Advanced Analytics on Engineered Feature to detect Sticker Breakout in Continuous Casters

Extra-Curricular

- Maintaining a literary blog featuring poetry, short stories and essays on various topics
- Team leader – Designing an efficient Recumbent tricycle. SAE-NIS: Efficycle.
- Worked as an executive member with the NGO Svagatagami in the project Abhilasha aiming to help the differently abled kids using technological innovation
- Green - one belt in Tae Kwon-do. Gold medal in district Taekwondo demonstration and Sparring
- Member of the University weight lifting team

- ***ML using sensor data*:** Sticker (rupture) formation in the surface of slab cast in the steel solidification process is a major concern. Temperature signals recorded by thermocouples embedded on the mold wall manifest specific signatures during such sticker formation. We used advanced machine learning techniques for pattern recognition of such time series signatures involving innovative feature engineering techniques. The solution is currently deployed in Tata Steel, boosting sticker detection accuracy and increasing production.

- ***Genetic Algorithm for vessel lineup optimization:*** demurrage charges payed by the ships at ports are complex non-linear rates that depend upon a host of factors. We developed a Genetic Algorithm based approach to arrive at a feasible vessel lineup that minimizes the demurrage costs.

- ***Social Media based Lead generation*:** Used NLP techniques to mine the content on social media sites like Facebook, Twitter, etc. to identify posts (and users) who may be interested in credit products (like cards, loans, etc.).

Master Thesis, IIT-Kanpur, India
Graduate Student *Mar 2015 – July 2016*

Key projects:

- ***A Two-Step Item Response Theory Model for India's Graduate Aptitude Test in Engineering (GATE)**:* Classical test scoring faces many problems. Fixed marks that represent difficulty leads to ties in the results. A post-facto analysis of the data gathered during the exam, the answer-sheets, can be used to dynamically determine the difficulty of the questions and award a unique rank to each student. We used Item Response Theory in conjunction with Maximum Likelihood Estimation to determine the unique ranks and difficulty. The results were correlated with the existing GATE score to ensure consistency in the predictions. The assumptions of Item Response Theory were challenged in the context of the Indian test taking scenario, such as the presence of the coaching industry and negative marking, and alternative formulations were proposed that better represent GATE. The same theory can be extended to design online test that dynamically learn the "quality" of the student and subsequently determines the best set of questions to gauge it.

- ***Parameter estimation of 3-parameter Fréchet Distribution**:* Find all parameters of a 3-parameter Fréchet distribution useful in describing the statistics from Non-Destructive Testing (NDT) of concrete using Ultrasonic Pulse Velocity (UPV) and find the confidence intervals for estimated parameter

GE India Innovation Centre (Aviation), Bangalore, India
Under-graduate intern *May 2014 – July 2014*

GE India Innovation Centre caters to the design, simulation and analytic needs of the General Electric. GE aviation has designed, developed and patented various turbofan and turboprop engines and related technologies.

Key projects:

- ***Prediction of Engine Failure**:* Used Cox's Proportional Hazard model on the available engine generated data to predict engine failure and flag the problematic engines sufficiently in advance to allow for servicing. Engineered features on vibration, pressure, and temperature data through taxi-takeoff-cruising-landing-braking to diagnose preventable ailments in the engine.
- ***Data downloading tool***: Developed a Java based software to facilitate the downloading of flight data through Oracle/Teradata based servers using a structured graphical interface. The tool was intended to make the database accessible for those who have limited knowledge of SQL. Automated the downloading process using a system scheduler to maintain an up-to-date local database.