

# Tirth Bhatt

## Data Scientist

✉ tirthc58@gmail.com ☎ 8735890054 🏠 Ahmedabad 🌐 [GitHub](#) [in](#) [LinkedIn](#)

### Profile

---

Data Scientist with **1 years** of experience comprising Machine learning, Data Analysis and Deep learning. Strong willingness to exhibit my proficiency in Deep learning and Machine Learning Algorithms in professional environment.

### Professional Experience

---

11/2021 – present  
Bangalore, India

#### Data Science Intern, *iNeuron.ai*

- Processing, cleansing and verifying the integrity of data used for analysis. Doing analysis and presenting results in a clear manner.
- Working on building and optimizing the state-of-the-art Machine learning and Deep learning models.
- Working and learning on how to create and automate the project lifecycle with the help of creating data pipelines.

04/2021 – 11/2021  
Vadodara, India

#### Trainee Data Analyst, *Rishabh Software*

- Applied Python programming skills, specializing in **pandas** and **NumPy**, for effective data manipulation and analysis.
- Utilized **AWS Athena** for SQL-like querying directly from **AWS S3**, showcasing competence in cloud-based data retrieval.
- Leveraged **Matplotlib** for creating impactful visualizations, enhancing data representation and interpretation.
- Developed insightful dashboards and reports using **Power BI**, translating complex data into visually compelling formats for data-driven decision-making.

### Projects

---

#### 1) Diamond Price Prediction, *iNeuron.ai*

- Successfully completed an End-to-End Machine Learning project on Diamond Price Prediction, featuring cloud deployment.
- Implemented **regression ML models** (Linear, Ridge, Lasso, Elastic Net, SVR, RF, Boosting) to predict diamond prices based on various features.
- Adopted **modular coding** for increased code reusability and a cleaner, maintainable codebase.
- Utilized **Git** and **Docker** for version control, ensuring smooth collaboration and efficient development
- Deployed the final application on both **AWS and Azure**, emphasizing accessibility for widespread usage.
- Employed a technology stack that included machine learning models for regression, Git for version control, Docker for containerization, **Flask as the framework**, and deployment on cloud platforms AWS and Azure.
- **Technology stack:** - **Machine Learning:** Regression (Linear, Ridge, Lasso, Elastic Net, svr, Random Forest, boosting) **Version Control:** Git **Containerization:** Docker **Framework:** Flask **Cloud Platforms:** AWS, Azure

## 2) Adult Census Income, *iNeuron.ai*

- **Problem Description:** Addressing an Income Prediction challenge with the Adult Income Census dataset, aiming to classify whether an individual earns more or less than \$50,000 annually.
- **Dataset Features:** Employing a **classification approach** in machine learning with features such as age, work class, education, marital status, and others, encapsulated in the 'Dataset columns' list.
- **Machine Learning Models:** Employing a diverse set of classifiers, including decision tree, random forest, support vector machine, bagging, and boosting, to achieve accurate predictions of income levels.
- **Workflow Tools:** Implementing **MLOps** and **Data Version Control (DVC)** for streamlined development, experimentation, and versioning of machine learning models.
- **Deployment Platforms:** Utilizing Docker and Git for containerization and version control, and deploying models on cloud platforms such as AWS and Azure for efficient and scalable predictions in real-world scenarios.
- **Technology stack:** - **Machine Learning:** SVC, Random Forest, Bagging, boosting **Version Control:** Git, DVC **Containerization:** Docker **Framework:** Flask **Cloud Platforms:** AWS, Azure **Mlops Tool:** Mlflow

## 3) Flight Price Prediction

- **Problem Description:** The objective of the study is to analyse the flight booking dataset obtained from “Ease My Trip” website and to conduct various statistical hypothesis tests in order to get meaningful information from it.
- **Dataset Features:** Employing a **Regression approach** in machine learning with features such as Airline, Source City, Destination City, Flight, Stops, Class, and derived features like Departure Time, Arrival Time, Days Left, and continuous feature like Duration.
- **Machine Learning Models:** Employing a diverse set of classifiers, including decision tree, random forest, support vector machine, bagging, and boosting, to achieve accurate predictions of income levels.
- Utilized **Git** and **Docker** for version control, ensuring smooth collaboration and efficient development
- **Workflow Tools:** Implementing **MLOps** and **Data Version Control (DVC)** for streamlined development, experimentation, and versioning of machine learning models.
- **Deployment Platforms:** Utilizing Docker and Git for containerization and version control, and deploying models on cloud platforms such as AWS and Azure for efficient and scalable predictions in real-world scenarios.
- **Technology stack:** - **Machine Learning:** SVC, Linear regression, Bagging, boosting **Version Control:** Git, DVC **Containerization:** Docker **Framework:** Flask **Cloud Platforms:** AWS, Azure **Mlops Tool:** Mlflow

## 4) Hotel Chain Dataset (Exploratory Data Analysis)

- **Problem Description:** The objective of the study is to analyse the Hotels industry dataset obtained from website and to conduct various test like Spatial Analysis and Descriptive to get meaningful information from it.
- **Descriptive Analysis:** Examined statistical measures to understand central tendencies and spread within the dataset.
- **Spatial Analysis:** Explored geographical aspects, potentially investigating hotel distribution and spatial patterns.
- **Choropleth Maps:** Visualized spatial data through Choropleth maps, offering insightful regional information.
- **Market Segment Analysis:** Investigated market segments to determine the segment with the highest number of bookings, providing valuable insights for the hospitality industry.
- **Python Libraries Used:** Leveraged Pandas, NumPy, Seaborn, Matplotlib, and Plotly for data manipulation, visualization, and analysis.

## Skills

### Programming languages:

Python

### Deep learning algorithms:

ANN, CNN, RNN

### Database

MySQL, MongoDB, Cassandra

### Cloud

AWS, Azure

## Education

### Machine learning algorithms:

Linear regression, Logistic regression, Decision tree, Random Forest, XGBoost, Bagging and ensemble

### Python Farmwork:

Flask, Rest Api

### Frameworks

Tensorflow, Sklearn

### MLops

Docker, mlflow, DVC

2021 – 2022  
Ahmedabad,  
India

**P.G.D in Data Science**, *Gujarat Technological University*

2017 – 2020  
Vadodara, India

**Bachelor's Degree**, *Gujarat Technological University*

## Certificates

- Having **Post Graduation** Diploma Certification in **Data science** with Distinction class.
- Completed **Amazone Web Services Foundations** badge certification.
- Completed **Full stack Data science Certification** with neuron.ai company certification.