

# Upendra Gosavi

## Data Scientist & Data Analyst

Upendrasgosavi@gmail.com

8689809107

Thane, Mumbai, Maharashtra

Linkedin

Github

Portfolio

"As a data professional, expertise in data science is a testament to strong analytical capabilities and problem-solving skills. Meticulous attention to detail is applied to navigate the intricate landscape of data, unlocking hidden potential. Let's embark on a collaborative journey to create data-powered solutions and shape the future of analytics collectively."

## Work History

### ISF Analytica and Informatica Pvt. Ltd.

April 2023- Sep 2023

Data Scientist

- Developed 5 interactive dashboards for the EasyLearn platform using Looker Studio and PowerBI.
- Created and optimized SQL views and queries, resulting in a 15% reduction in data extraction time.
- Conducted data analysis on a dataset of 50,000 records, identifying key trends and insights.
- Collaborated with a team of 3 data scientists, contributing to 2 successful data-driven projects.
- Pre-processed and cleaned a dataset of 10,000 records, ensuring data quality and accuracy.

### Ai Variant

Sep 2022 - Feb 2023

Data Scientist (intern)

- Pre-processed and cleaned data for analysis, handling a dataset of 20,000 records.
- Analyzed data using statistical techniques, leading to valuable insights.
- Completed a comprehensive data science internship program, participating in 6 projects.
- Collaborated with a team of 5 data scientists and received mentorship from experienced professionals.

## Technical Skills

- Languages : **Python (NumPy, Pandas, Scikit-learn, Matplotlib)**
- Databases : **MySQL, SQL Server.**
- Data Visualization : **Power BI, Looker Studio, Seaborn, Tableau**
- Data Wrangling : **Web Scraping using BeautifulSoup , Data Mining, Data Preprocessing**
- Machine Learning : **Regression, Clustering, Classification, Recommendation System, Natural Language Processing (NLP)**
- Deep Learning : **CNN, ANN, Keras ,TensorFlow**
- Microsoft Office : **Advance Excel, LOOKUP, Pivot Table, VBA Excel, HLOOKUP, MIS Reporting, Macros, Powerpoint**
- Data Modeling : **SQL, ETL Tools**
- Data Science : **Supervised Learning, Unsupervised Learning**

## Skills And Interests

- Goal Setting, Knowledge Acquisition
- Details Oriented, Data Integrity
- Reporting , Analysis Skills,
- Strategic, Attention to Detail,
- Statistical Analysis,
- Data Professional
- Presentation skills
- Key Performance Indicators
- Data Analytics, Data-driven Decision Making

## Certifications

**Certified Data Scientist**  
ExcelR - 2023

March 2023

**IBM Certification in Data Science**  
2023

June 2023

**Mastars Program In Data Science**  
NASSCOM - 2023

April 2023

**Python for Data Science, AI & Development**  
Data Science Corner - 2022

October 2022

## Education

**SSC**

March 2014

**Mumbai University**

April 2019

Diploma In Information technology  
C, C++, Java , Python

### Customer Churn

- Data collection and preprocessing: Collect relevant data from various sources, clean and preprocess the data. Handle missing values and encode categorical variables.
- Exploratory data analysis (EDA): Perform EDA to understand the data and relationships between features and the target variable.
- Feature engineering: Create new features that can be relevant to the problem statement. Model selection: Choose the appropriate algorithm(s) for the problem. This may include supervised learning algorithms such as logistic regression, decision trees, random forest, or neural networks.
- Model training: Split the data into training and validation sets, train the model using the training set, and tune the model parameters to achieve optimal performance.
- Model evaluation: Evaluate the performance of the model on the validation set. Use appropriate metrics such as accuracy, precision, recall, or F1 score.
- Model deployment: Deploy the trained model in a production environment for real-world predictions.

### Leaf Disease Detector

- Data collection: Gathered a comprehensive dataset comprising images of both healthy and diseased leaves from various sources, including field photographs.
- Data preprocessing: Preprocess the data by resizing the images, normalizing pixel values, and splitting the data into training, validation, and testing sets.
- Model architecture selection: Choose the appropriate CNN architecture for the problem, such as VGG, ResNet, or Inception.
- Model training: Train the model using the training set, using techniques such as data augmentation, transfer learning, or fine-tuning.
- Model evaluation: Evaluate the performance of the model on the validation and testing sets using metrics such as accuracy, precision, recall, or F1 score.
- Model deployment: Deploy the trained model in a production environment to detect leaf diseases automatically.

### Netflix TV Shows and Movies Analysis using Tableau

**Analyzed Netflix's content library to understand the distribution of TV shows and movies, with a focus on various aspects of the content.**

- Utilized bar plots to illustrate the number of shows falling under different rating categories (e.g., IMDb ratings).
- Employed maps to pinpoint countries with the highest representation in Netflix's content library for both TV shows and movies.
- Explored the distribution of genres across Netflix's content, providing insights into the most popular genres.
- Visualized the historical evolution of TV shows and movies on Netflix by the release year, highlighting trends and growth.
- Identified and showcased the top 10 most prevalent genres within Netflix's content library.
- Gained valuable insights into Netflix's content strategy and user preferences, which can inform content acquisition.
- Visit Dashboard to explore this project and other data analysis work.

**Dashboard Link**

### TripAdvisor Insights: Unveiling Hospitality Dynamics

**Key Stats Overview: Compiled total hotels, average rooms, user reviews, and scores for an industry snapshot.**

- **User Engagement:** Visualized user activity by time period through a dynamic bubble chart.
- **Hotel Ratings Distribution:** Bar chart showcased hotels by star ratings, indicating industry quality distribution.
- **Travel Preferences:** Used treemaps to reveal prevalent travel types, highlighting user preferences.
- **Global User Demographics:** Horizontal bar chart displayed user distribution across continents.
- **Top Hotels Highlight:** Identified top 10 hotels by total room bookings for spotlighting popular accommodations.
- **User Service Preferences:** Analyzed additional services preferences via a concise text table.

**Dashboard Link**