

### SUMMARY

Highly motivated Junior Data Scientist with more than one year of practical experience in data analysis, machine learning, statistical modeling, creating and deploying APIs. Proficient in extracting actionable insights from complex datasets through data preprocessing, visualization, and predictive modeling. A strong problem solver and communicator, good at collaborating with multidisciplinary teams, and conveying technical findings to non-technical stakeholders. Eager to apply my skills and passion for data-driven solutions to drive organizational success.

### PROFESSIONAL EXPERIENCE

#### Junior Data Scientist

Jul '22 - Present

#### Onpassive Technologies Pvt Ltd.

Hyderabad, IN

*Product Based AI Technology Company developing fully autonomous Software as a Service (SaaS) products for global clientele.*

#### Projects

- Worked on various projects like Speech translation modules, Sentence Correction and Smart Compose feature in Grammar check using state-of-the art pre-trained transformers, and Image to document conversion - Text extraction using OCR by training object detection module using YOLOv5. Currently working on EKYC feature which involves face recognition, Liveness detection, and OCR modules

#### Overview

- More than 1 year of experience in Machine Learning, Data Science with large datasets of Structured and Unstructured data, Data Acquisition, Data Validation, Data pipelines, Predictive Modeling, Data Visualization, Testing and Deployment.
- Working knowledge of Python for data extraction and data manipulation, and widely used python libraries like Scikit-Learn, NumPy, Pandas, NLTK, Matplotlib, Seaborn, SciPy, statsmodels.
- Extensive Research on architecture for the above POCs and features and developed end-to-end architecture for Speech Translation and Image to Document Conversion POCs
- Good knowledge in creation and retrieval of Data from databases like SQL, MySQL and MongoDB
- Good knowledge of implementing machine learning algorithms and proficient with developing neural networks such as CNNs, RNNs, attention and Transformer Models using TensorFlow and PyTorch.
- Gained knowledge on CI/CD pipelines for efficient and automated model deployment.
- Hands-on experience in creating APIs using FlaskAPI, FastAPI.

#### Data Science Intern

Jan '22 - Apr '22

#### Onpassive Technologies Pvt Ltd.

Hyderabad, IN

*Product Based AI Technology Company developing fully autonomous Software as a Service (SaaS) products for global clientele.*

#### Projects

- Email Category Classification** - This feature is used to classify an email into one of the following categories - Social, Primary, Updates and Promotions.

#### Overview

- Worked extensively on cleaning and analyzing around 3 lakh rows of email data to build a classification model to predict the category of an email
- Developed a classification model using Naive Bayes classifier to classify the category of an email
- Designing Streamlit application to demonstrate the project to stakeholders

## EDUCATION

<b>MSc. in Data Science</b>	Oct '20 - Jul '22
<b>GITAM University</b>	Hyderabad, IN
CGPA - 8.89	
Key Courses - Python, Machine Learning, Statistics, SQL, NoSQL, Cloud Computing	
<b>BSc. in Computer Science</b>	Jun '17 - Sep '20
<b>Bhavan's Vivekananda College</b>	Hyderabad, IN
CGPA - 8.81	
Specialization- Mathematics, Statistics and Computer Science	

## DATA SCIENCE PROJECTS

- **Master Thesis - Malicious URL Detection | Tech Stack: Python | March '22**
  - The aim of this project is to classify a URL as a Malicious or Benign URL. Collected a public dataset that consists of 450176 rows and two classes of URLs: Malicious or Benign. Have extracted a total of 17 features which consist of Lexical features, count-based features, and two binary features.
  - Trained the model on AdaBoost and Random Forest classifiers and used the Voting Classifier ensemble method to get the optimum result among the two classifiers.
- **Exploratory Data Analysis - Loan Default Dataset | Tech Stack: Python | July '22**
  - Performed EDA on Lending Club Dataset which consists of around 111 columns and around 39k rows of unique customer details. Cleaned and analyzed the data using different techniques to derive insights from the data.
  - Performed Univariate, Segmented Univariate and Multivariate Analysis to gain insights and relationships between the variables.

## KEY SKILLS

• Data Analysis	• Machine Learning	• Deep Learning	• Statistical Modelling
• Programming	• Data Visualization	• Data Wrangling	• Databases
• Version Control	• Statistical Methods	• Data Manipulation	• Modeling Techniques
• Data Visualization Techniques	• Natural Language Processing	• Quantitative Analysis	• Debugging

## TECHNICAL SKILLS

<b>Programming Languages</b> : Python, C, C++, Java, HTML
<b>Python-toolboxes</b> : NumPy, Pandas, Scikit-learn, Tensorflow, PyTorch, Keras, Matplotlib, OCR, Easy-OCR, Seaborn, Plotly, Statsmodels
<b>Data Visualization Tools</b> : Microsoft Excel, Power BI
<b>Databases</b> : SQL, NOSQL, MONGODB
<b>IDEs</b> : Android Studio, Anaconda, Jupyter Notebook, Google Colab
<b>Version Control</b> : Git - GitHub, Bitbucket
<b>Additional</b> : Streamlit, FastAPI, FlaskAPI, Docker, AWS S3 storage Service, CI/CD Pipeline, Latex, Microsoft Office Suite

## CERTIFICATIONS

- Advanced Certification Program- ML and DL | IIIT Bangalore | Jul '23 - Present
- Android Development —Android Oreo Developer Course| Udemy | Dec '18

## ADDITIONAL ACHIEVEMENTS & VOLUNTEER EXPERIENCE

- Received second prize for my android Application - Onestop at a College-level Tech competition during my Bachelor's.
- Honored with "Mr. Computer genius 2019" award for my achievements and projects in Computer Science.
- The app which I have developed "Bhavan's Vivekananda College App" is being officially used by the college with over 2000+ users.
- Worked as Vice President of Bhavan's Core Division in an organization called Street Cause which is the largest student-run NGO.