

Abhishek Venugopal

Data Analyst

abhishekvpta@gmail.com

Pathanamthitta

+91 8075364365

[LinkedIn](#)

[GitHub](#)

[Medium](#)

EDUCATION

Degree	Institution	GPA/ Percentage	Year
BS-MS in Biological Sciences	IIT Madras, Chennai	7.22 /10	2020(Dual degree)
Class 12 Certificate (CISCE)	Evershine School, Pathanamthitta	92%	2015
Class 10 Certificate (CISCE)	KE School, Kottayam	93.4%	2013

COURSES AND CERTIFICATIONS

Data Structures and Algorithms	Advanced Python	SQL	Bioinformatics
Data Visualizations with Plotly	Biostatistics	Pandas	Power BI dashboard

SKILLS

Languages: Python, C, Java, SQL	Data analysis: NumPy, SciPy, Pandas, StatsModels
Data visualization: Matplotlib, seaborn, Plotly	Deployment tools: Flask, PyWebIO, Heroku cloud
Machine Learning: scikit-learn, PyTorch, GBMs	Tools: Jupyter, VS Code, Power BI, Tableau, MySQL

EXPERIENCE

Data Science Intern | Papertown

May 2021 – present

- Selected as an intern in data science group based on profile and experience among **200+** applicants through LinkedIn.
- Effectively contributing as a team member on competitions with real-world data, conducting extensive **exploratory data analysis** and implementing various **machine learning algorithms** and its **deployment**.

Super coordinator WebOps | Biotech Student's Association (BSA) | IIT Madras

Jan 2019 – March 2019

- Setup the automated registration **data collection** and effectively handled the communication to all participants of Biofest'19 which saw a footfall of **400+** participants using **MySQL** and **Excel** spreadsheets.
- Coordinated two **technical** workshops on Computational biology and Protein Visualization which saw a footfall of **100+** attendees.

Summer Research Intern | IISc Bangalore: Molecular Biology

April 2019 – May 2019

Summer Research Intern | IISER Thiruvananthapuram: Structural Biology

April 2017 – May 2017

PERSONAL PROJECTS

1) Exploratory data analysis of COVID-19 in India

[Link to project](#)

- Scrapped the **dynamic COVID-19 medical information** of State/Union Territories page from Wikipedia. Rendered an interactive zoom-able map using **Folium library** along with insightful **visualizations** of the statistics of the pandemic using **seaborn**.

2) Sales analysis of electronic goods using Pandas

[Link to project](#)

- Extracted monthly sales data of electronic items in US states in 2019 (from GitHub), combined them and derived useful **business insights** and answered relevant questions which can translate to improved business through **data-driven analysis** by extensively using **Pandas data analysis and mining** techniques.

3) Login system using python tkinter with connectivity to MySQL

[Link to project](#)

- Created a functional **GUI based employee login/registration system** using **tkinter** and **MySQL connectivity** through python for authentication, with appropriate display of error messages with **regex**.
- Applied **object oriented programming** for creating two classes along with buttons and trigger functions.

4) E-commerce products classification using LightGBM and OPTUNA

[Link to project](#)

- Conducted **EDA** and **feature selection** in a highly imbalanced dataset containing **0.1M** records and **48** right skewed categorical features and four classes.
- Experimented with **SMOTE**, data transformation and outlier detection techniques and prepared the data for **Random Forest, XGBoost and LightGBM pipeline** using OPTUNA hyper-parameter optimization with 20 trials.
- Achieved the best **log loss** (evaluation metric) score of **1.086** with LightGBM model, placing it in the **top 25 %** of all predicted models in Kaggle.