Abhishek **Venugopal**

abhishekvpta@gmail.com   **Pathanamthitta**   +91 8075364365   **LinkedIn**   **GitHub**   **Medium**

EDUCATION

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

RELEVANT COURSES & SKILLS

Data Structures and Algorithms	Advanced Python	SQL	Bioinformatics
Data Visualizations with Plotly	Biostatistics	Pandas	Power BI dashboard
C Programming fundamentals	Differential calculus	Storytelling	Critical thinking

Languages: Python, C, Java, SQL   **Data analysis**: NumPy, SciPy, Pandas, Excel

Data visualization: Matplotlib, Seaborn, Plotly   **Deployment tools**: Flask, streamlit, GitHub, Heroku

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 team based on profile among **200+** applicants through LinkedIn.
- An Edtech startup aimed at bringing together similar learning communities riding the booming **creator economy**
- Conducting **statistical** and **time series** analysis and implementing various **machine learning algorithms**

Super coordinator WebOps | Biofest | 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 from various colleges

Summer Research Intern | IISc Bangalore: Molecular Biophysics   March 2019 – May 2019

Summer Research Intern | IISER Thiruvananthapuram: Structural Biology   March 2017 – May 2017

PROJECTS

Resume Optimizer Web app – NLTK, Streamlit, Heroku   Link to application

- Created and deployed my own resume optimizer app from scratch, **used frequency, tagging ,wordcloud in NLP**
- Employed a **self-created resume scoring technique** along ATS guidelines with bonus and penalty for keywords
- Added an experimental salary prediction feature, used **multiple regression, conditional weights and biases**
- Next update is aimed at creating a top keywords **Excel database**, separating hard and soft skills for each job title

Real estate property search tool - Machine Learning regression, streamlit, Heroku   Link to application

- Deployed a **web app** which predicts the **price** of a property based on user choices and its **location on the map**
- Used a pipeline of machine learning **regression models,** chose least **RMSE,** mapped coordinates using zip codes

Sales analysis of electronic goods – NumPy, Pandas, matplolib, seaborn   Link to notebook

- Extracted monthly sales data of electronic products 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 an extensive **data-driven analysis**

World Happiness Report 2021– Pandas, seaborn, Plotly chart studio   Link to article

- Conducted an **exploratory data analysis** to find out which countries are happier in 2021 and why or why not
- Created highly **interactive** visualizations using **Plotly**, published an article about the findings on Medium

E-commerce products multiclass classifier – sklearn, LightGBM, OPTUNA   Link to notebook

- 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 used **OPTUNA** hyper-parameter optimization, achieving 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

Exploratory data analysis of COVID-19 – NumPy, Pandas, Folium   Link to notebook

- Extracted the **dynamic** COVID-19 medical information of State/Union Territories in India page from Wikipedia.
- **Cleaned**, analyzed and rendered an interactive zoom-able map using **Folium** library along with insightful visualizations of the statistics of the pandemic using **seaborn**

ACHIEVEMENTS & ACTIVITIES

- Served as the **captain** of Hostel Chess Team winning **Gold** at Schroeter, Institute level tournament   (2020)
- Recipient of Half-Time Teaching Assistantship **Scholarship** (HTTA) awarded to GATE qualified students   (2019)
- **Volunteered** at National Service Scheme(NSS), IIT Madras, involved in drives and projects   (2015-2016)
- **Perfect attendance** record throughout School, best outgoing student by academic performance   (2013)