GOPALAKRISHNAN KUMAR
A-502 , Sai Nidhi CHS , Opp Chembur Karnatka High School
ChemburMumbai-400071
Email - gopal.krishnan.18.1.1@gmail.com
Linkedin Profile :https://www.linkedin.com/in/gopalakrishnan-kumar-a73301110/
Mob +91-8928316507 / 9819043881.

An IIT Bombay alumnus (M.Tech in Chemical engineering) with a robust background in Systems Biology and over 6 years of hands-on experience in the dynamic field of data analytics.

🔍 What I Offer:

- Data cleaning, analysis, and visualization

- Advanced statistical modeling and predictive analytics

- Expertise in Python, R, SQL, Tableau, Excel, and more

- AI and ML techniques including deep learning and NLP

🏆 Certifications & Achievements:

- Certified by IBM, Harvard, MIT, and Microsoft , IIT- kanpur , Intellipaat

- Recognized as a Kaggle Competition Expert

-Secured All India Rank 41 in the Gate exam.

- Published research work during a prestigious scholarship in Paris

💼 Why Choose Me:

- Proven track record of delivering high-quality results on time

- Ability to solve complex problems and translate data into actionable insights

- Commitment to effective communication and collaboration

SKILLS

- **Operating System -**Windows, Linux**-** Ubuntu
- **Data Science**- Data analysis, R statistical computing, connecting R with Hadoop framework, Machine Learning algorithms, time-series analysis, K-Means Clustering, Naïve Bayes, business analytics
- Explanatory Data Analysis -Data cleaning and pre-processing to ensure data quality and integrity.
- **Machine Learning-** Python, Algorithms, Statistics & Probability, Supervised & Unsupervised Learning,Decision Trees, Random Forests, Linear & Logistic regression, Arima Model
- **Machine Learning Model Engineering Model Selection**
    Kaplan-Meier Estimator
- **AI-** convolutional neural networks (CNN), perceptron in CNN, Tensor Flow code, transfer learning, graph visualization, recurrent neural networks (RNN), Deep Learning libraries, GPU in Deep Learning, Keras
- **ChatGpt** Writing of effective Prompts,Integration of Chat GPT with MS Excel Powerpoint
- **Python-** Anaconda, Lambda expression, OOP, NumPy, SciPy, MatPlotLib, JSON, Packages, Functions,Web scraping, Python parser ,Skikit Learn **HTML, Javascrit, CSS**
- **Tableau-** Creating Tableau Dashboards,Tableau Desktop and public integration with R and Big Data
- **Power BI –**Data Visualisation
- **SQL Server -** SQL architecture, client/server relation and databases, Subqueries, Transact-SQL ,DatabaseObjects, Implementing Programmability Objects, Database Concurrency ,SQL Lite,

- **Big data Hadoop-** MapReduce, HDFS, Advanced Hive and Impala, Pig, Sqoop, Oozie, Flume and HBase Spark framework and RDD, Scala and Spark SQL, Machine Learning using Spark
- **Matlab ,Scilab, -** Image processing
- **Fortran, Clinical SAS Programming**
- **Google Analytics Application**
- **Cyber Security :** Securing Data from potential frauds using Python

----

Projects Completed in Data Science

Projects using Python and R
Project 1 Innodatatics, Hyderabad (Aug 2020 to Nov 2020)
Project Title : Identifying Risk factors influencing Diabetes Type – Classification Model.
Domain : Health Care
Roles and Responsibilities : **Junior Data Analyst- Team Leader**
Problem Statement
Collection and analysis of data received from the diagnostic department which had recorded data on a few regular patients who visit hospital more often. The dataset consisted of data recorded from 2018 to 2019. The dataset focuses on patients suffering from "Type 1 diabetes", "Type 2 diabetes", and a few other diseases. The aim is to predict and cluster the patients as per the types of diabetes they suffer and admit them in the proper ward. Building a machine learning model to group them accordingly to clear up the confusion that is being created in the hospital. The purpose of this project is to predict the ability of machine learning for classifying the types of Diabetes Mellitus (DM) using factors that determine diabetic types in susceptible individual, thus helps health care providers in different settings to find out the disease in their earliest phase which in turn enhance good prognosis which will bring good business benefits.

Technology Stack and Algorithms used :
Python : Machine Learning
Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used : In this project, Unsupervised learning namely K Means clustering and Gradient Boosting Regression algorithm are used for obtaining high accuracy percentage. Also Supervised learning algorithms namely Neural Network, LR, LDA, KNN, CART, NB, and SVM, is used to effectively verify Model accuracy of the models.
Steps Used: Descriptive Analysis and Predictive Analysis and for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and, evaluation of models are done using Python and R.

Project 2- Innodatatics , Hyderabad (Dec 2020 to May 2021)
Project Title:
Data Exploration and Forecasting on Patient Count – Regression Model
Domain : Health Care
Roles and Responsibilities : **Junior Data Analyst- Team Leader**
Problem Statement
The purpose of the project is to predict the ability of Machine Learning (ML) techniques to analyze the data pertaining to Out-patients visiting different hospitals with respect to year-wise, age-wise, stage-wise, specialty-wise,time-bound. Our approach involves descriptive analysis and processing of data, building machine learning models and evaluating their accuracies, identifying challenges and constraints in the execution of the projectwhich will bring good business benefits.
Technology Stack and Algorithms used :
Python : Machine Learning
Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used : Unsupervised Learning Algorithms namely Arima() Model, Auto Regressor() Model,Stationarity Analysis.
Steps Used: Descriptive Analysis and Predictive Analysis and Time Series Analysis for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and evaluation of models are done using Python and R.

Libraries used: Pandas, Scikit-learn, Numpy, Keras, Tensor flow

----

Project 3 Innodatatics , Hyderabad (July 2021 to Sept 2021)
Project Title :Infant Mortality Rate Analysis –Regression Model
Roles and Responsibilities : **Junior Data Analyst- Team Leader**
Problem Statement

To find the proportion of infant death registrations according to their respective GEO information To visualize each age category at the time of death of an infant to understand their death rate ,To predict/forecast the total no. of deaths for every age group listed in the data for the next 5 years.

Technology Stack and Algorithms used :
Python : Machine Learning
Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used : Unsupervised Learning Algorithms namely Arima() Model, Auto Regressor() Model,Stationarity Analysis.
Steps Used: Descriptive Analysis and Predictive Analysis and Time Series Analysis for all steps of variable description, exploratory data analysis, data processing and preprocessing, feature selection, model building, and,evaluation of models are done using Python and R.
Libraries used: Pandas, Scikit-learn, Numpy, Keras, Tensor flow,

----

Project 4 Innodatatics , Hyderabad (Oct 2021 to Dec 2021)
Project Title :Survival Analysis – Unemployment Index–Classification Model
Domain : General -Social
Roles and Responsibilities: **Junior Data Analyst**
Problem Statement
•   To describe the unemployment index of survival group with respect to spell using Supervised Machine Learning method.
•   To describe the unemployment index of unemployment group with respect to spell using Supervised Machine Learning method.
•   To classify the unemployment index of survival and unemployment group with respect to timeline estimate using Supervised Machine Learning method.
•   To apply suitable machine learning model for classifying the unemployment index with respect to spell and event
•   To create awareness of unemployment rate so as to suggest ways and means to minimize unemployment rate
       Constraints:
The dataset has very few variables. Limited open source and cross platform to study the dataset. A possibility of model under fitting due to low variance

Project Pipeline : Data Understanding,Data Cleaning ,Machine Learning Model Engineering
Model Selection-Kaplan-Meier Estimator ,Model Testing and Evaluation,Model Deployment
Monitoring and Maintenance
.
Technical Stacks and Algorithms used in this Project
Languages: Python, html, cloud, sql, heroku, R, visual studio, git bash
AI/ML: Pytorch, skikit learn,
Libraries: lifelines, pandas, numpy, skikit learn
Database: Jupyter Online
Warehouse: Jupyter classic notebook
ETL: Python, Jupyter
Visualizations: Plotly()
Tracking & SC: Github
 Deployment Strategy:
 Web Based- HTML applications
 Enterprise Based- AI algorithm/ Kaplan Meier Fitter Model

Project 5 - Intellipaat Bengaluru
Domain: Banking
Project Description Loan Eligibility Prediction – Classification Model
Technology Stack and Algorithms used :
Python : Machine Learning
Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used: In this project, Unsupervised learning namely K Means clustering and Gradient Boosting Regression algorithm are used for obtaining high accuracy percentage. Also Supervised learning algorithms namely Neural Network, LR, LDA, KNN, CART, NB, and SVM, is used to effectively verify Model accuracy of the models.

Projects 6-Intellipaat ,Bengaluru using Tableau Desktop
Project : Covid-19 mortality rates.
Domain: Healthcare
Roles and Reaponsibilities Project Intern in Data Science Architect Masters Program
Project Description:
Based on the csv.files for global mortality rates, we have analyzed and developed a dashboard to understand the covid-19 global cases. Using filters, parameters and actions wherever possible to make the dashboard interactive.Comparing the global confirmed vs. death cases in a world map using pie charts.Creating a parameter for percentile comparison between countries based on confirmed or death cases. Comparing the country wise cases using logarithmic axes. Dashboard displaying both log axis chart and a default axis chart in an alternate interactive way. New cases per day in China and India – compared in a date wise chart.Which day has the highest new death cases?
Average WHO region wise cumulative cases visualized using a funnel chart. Dashboard created with a drop down menu to view the WHO region wise data using a bar chart, line chart or a map as per user's requirement.

Project 7 – Intellipaat ,Bengaluru using Microsoft SQL
Project Domain: Sales Analysis
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description:
Querying a large relational database using Adventure work database with SQL server with respect to.
- ⬜details from the personal table including email ID, phone number, and phone number type
- details of the sales header order
- ⬜details of the sales details order made in a month
- ⬜total sales made in a month
- ⬜total sales made in the year by month order by increasing sales
- ⬜total sales made to the customers

Project 8 - Intellipaat ,Bengaluru
Project Domain: Travel
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Big Data Hadoop Spark
Project Description:
Analysis of Taxi App pertaining to number of trips,no of passengers travelled, total number of kilometres travelled, rate per trip and revenue generated.

Project 9- Intellipaat ,Bengaluru
Project Domain: E **commerce website**
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program

Big Data Hadoop
Project Description
Analysis of the public review of the products on the Social media
Java, eclipse, neon, javascript, java compiler, apache

Project 10- Intellipaat ,Bengaluru
Domain :Social Media
Project Domain: E commerce website
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description
Analysis of Cosmetic Products on Social Media Platform with
respect to likes, dislikes
Technology used :Spark

Project 11Intellipaat , Bengaluru
Roles and Responsibilities : Project Intern Data Science Madder's Architect Program
Project Description
Title : Binary Classification on Customer Churning using Keras
Analysis of reasond why customers are switching to other competitors
Domain : Telecom Industry
Technology used :Artificial Intelligence keras

Project no 12 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :Bank Customer Data
Project Description
The objective of this project is to understand the UK bank customer data
Based on the .csv file provided analysis was done and developed a dashboard to
understand the customer data of a UK bank. Use filters, parameters and actions
wherever possible to make the dashboard interactive
Technology used :Tableau desktop
Interactive dashboard wasused to understand the data. Multiple features, various charts
suitable to the respective problem statements are to be used at points where chart type is not
specified

Project No 13 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :E commerce
Project Description :
Building an AI-based Chatbot using IBM Watson LAB
Domain :E Commerce
Technology used : Artificial Intelligence & Deep Learning with Tensorflow
----

Project No 14 – Intellipaat , Bengaluru
Roles and Reaponsibilities : Project Intern Data Science Madder's Architect Program
Project Domain :Analytics
Project Description
K-Means cluster analysis on Iris dataset to predict the class of a flower using its petal's
Dimensions .Using the famous Iris dataset, predict the
class of a flower Perform K-Means cluster analysis
Technology used :SAS

International Data Science   Kaggle Project Competitions
Kaggle Status: Notebook Expert / Competitions Contributor
Kaggle Profile : https://www.kaggle.com/gopalkk1
Winner of 10 bronze medals

1 Image Matching Challenge 2023
Reconstruct 3D scenes from 2D imagesResearch · Code Competition · 494 Teams · 8 months ago

2 Lux AI Season 2
Terraform Mars!Featured · Simulation Competition · 646 Teams · 10 months ago

3 Google - Fast or Slow? Predict AI Model Runtime
Predict how fast an AI model runsResearch · 616 Teams · 3 months ago

4 Benetech - Making Graphs Accessible
Use ML to create tabular data from graphsFeatured · Code Competition · 608 Teams · 8 months ago

5 Stanford Ribonanza RNA Folding
Create a model that predicts the structures of any RNA moleculeResearch · 755 Teams · 3 months ago

6 Bengali.AI Speech Recognition
Recognize Bengali speech from out-of-distribution audio recordingsResearch · Code Competition · 744 Teams · 4 months ago

7 RSNA 2023 Abdominal Trauma Detection
Detect and classify traumatic abdominal injuriesFeatured · Code Competition · 1125 Teams · 4 months ago

8 1st and Future - Player Contact Detection
Detect Player Contacts from Sensor and Video DataFeatured · Code Competition · 939 Teams · a year ago

9 Google Research - Identify Contrails to Reduce Global Warming
Train ML models to identify contrails in satellite images and help prevent their formationResearch · Code Competition · 954 Teams · 6 months ago

10 Santa 2023 - The Polytope Permutation Puzzle
Solve twisty puzzles in the fewest movesFeatured · 1054 Teams · 24 days ago

11 Open Problems – Single-Cell Perturbations
Predict how small molecules change gene expression in different cell typesFeatured · 1097 Teams · 3 months ago

12 SenNet + HOA - Hacking the Human Vasculature in 3D
Segment vasculature in 3D scans of human kidneyResearch · Code Competition · 1149 Teams · 18 days ago

13 AMP®-Parkinson's Disease Progression Prediction
Use protein and peptide data measurements from Parkinson's Disease patients to predict progression of the disease.Featured · Code Competition · 1805 Teams · 9 months ago

14 BirdCLEF 2023
Identify bird calls in soundscapesResearch · Code Competition · 1189 Teams · 9 months ago

15 Stable Diffusion - Image to Prompts
Deduce the prompts that generated our "highly detailed, sharp focus, illustration, 3d renders of majestic, epic" imagesFeatured · Code Competition · 1231 Teams · 9 months ago

16 AI Village Capture the Flag @ DEFCON31
Collect flags by evading, poisoning, stealing, and fooling AI/MLFeatured · 1344 Teams · 3 months ago

17 Vesuvius Challenge - Ink Detection
Resurrect an ancient library from the ashes of a volcanoFeatured · Code Competition · 1249 Teams · 8 months ago

18 UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN)
Navigating Ovarian Cancer: Unveiling Common Histotypes and Unearthing Rare VariantsResearch · Code Competition · 1326 Teams · 2 months ago

19 Parkinson's Freezing of Gait Prediction
Event detection from wearable sensor dataResearch · Code Competition · 1379 Teams · 9 months ago

20 Novozymes Enzyme Stability Prediction
Help identify the thermostable mutations in enzymesFeatured · 2482 Teams · a year ago

21 Child Mind Institute - Detect Sleep States
Detect sleep onset and wake from wrist-worn accelerometer dataFeatured · Code Competition 1877 Teams 3 months ago

22 Linking Writing Processes to Writing Quality
Use typing behavior to predict essay qualityFeatured · Code Competition · 1876 Teams · a month ago

23 Emoji_events
CommonLit - Evaluate Student Summaries
Automatically assess summaries written by students in grades 3-12Featured Code Competition · 2064 Teams 3months ago

24 Kaggle - LLM Science Exam
Use LLMs to answer difficult science questionsFeatured · Code Competition · 2664 Teams · 4 months ago

25 GoDaddy - Microbusiness Density Forecasting
Forecast Next Month's Microbusiness DensityFeatured · 3547 Teams · 8 months ago

26 ICR - Identifying Age-Related Conditions
Use Machine Learning to detect conditions with measurements of anonymous characteristicsFeatured · Code Competition · 6430 Teams · 6 months ago

27 LLM - Detect AI Generated Text
Identify which essay was written by a large language modelFeatured · Code Competition · 4358 Teams · a month ago

28 American Express - Default Prediction
Predict if a customer will default in the futureFeatured · 4874 Teams · 2 years ago

----

Projects done in training program with IIT Kanpur – Jan 2023- Sept 2023

DATA SCIENCE PROJECTS DONE DURING DATA SCIENCE TRAINING PROGRAMS–Edvancer-IIT Kanpur

Machine Learning Projects- Python
Consumer Services- Consumer Complaints Resolutions
- ✓ BFSI Marketing- Understanding Customer Preferences in Insurance Sector
- ✓ Pharma Public Safety- Counterfeit Medicines- Prediction of Sales
- ✓ Manufacturing- Predict Hazard Ratings for a Maintenance Project
- ✓ Real Estate-Flag Junk Property Listings
- ✓ Health Care- To Predict No-shows given the appointment details

Artificial Intelligence Projects:
- ✓ Multiclass Multilabel prediction For stack overflow- Given text for Questions , predict tags associated with them
- ✓ Music Genre Identification- Given audio files for songs , identify which genre they fall in
- ✓ Spam filter for Quora questions
- ✓ Distracted Driver Multi Action Classification Page-Classification of the various distractions of a driver
- ✓ Image Captioning Page- Uploading photos and pixel analysis

----

Project Proposal Submitted in Freelncer Upwork Site –Jan 24- Feb 24 and Ongoi

1 Proposal for Project: Advanced Algebraic Equation Solving and Exponential Regression

Steps Used
Advanced Algebraic Equation Solving
Exponential Regression
2.Proposal for Project : Health Care Data Scientist .
Develop predictive models to improve patient outcomes and optimize healthcare operations.

- Analyze clinical data to identify trends, patterns, and potential areas for intervention.

- Implement machine learning algorithms for disease prediction, diagnosis, and treatment recommendation.

- Collaborate with cross-functional teams to design and deploy scalable data-driven solutions.

3 Proposal for Project: Develop R Shiny App

The objective of this project is to develop a user-friendly web application using R Shiny that meets your specific requirements. Whether you need a dashboard for data visualization, a tool for interactive analysis, or a platform for sharing insights, I am committed to delivering a high-quality solution tailored to your needs

----

Academic Projects-Computational Biology

1. **At IIT Bombay–Cell-to-Cell variability during TNFa Signaling and during Yeast Cell Cycle [M.Tech**
Project, Guide - Prof. Ganesh A. Viswanathan]

2.**At INRIA Paris (Rocquencourt, France) - Stochastic Modeling of the Yeast Met3 Promoter [Internship**
Project, Guide - Prof. Gregory Batt, Apr 2011- June 2011]
Skills used : MATLAB, SCILAB, Fortran

Research Publications – Systems Biology- M.Tech Course at IIT Bombay
Dhananjaneyulu V, Sagar, P VN, **Kumar G**, Viswanathan GA (2012): Noise Propagation in Two- Step Series MAPK Cascade. PLoS ONE 7(5): e35958. doi: 10.1371/journal.pone.0035958 [Cited 13 Times]
Link : **https://doi.org/10.1371/journal.pone.0035958**
Skills used- MATLAB, SCILAB, Fortran

WORK EXPERIENCE

1.Presently a Free-lance Data Science Consultant with GLV Data Solutions, Mumbai from Nov 2023 tll date.

Project Description :
i).Nutrition, Physical Activity, and Obesity - Behavioral Risk Factor Surveillance System On Going –from Feb 2023
Project Domain : Nutrition and Health Care
Roles and responsibilities : Data Scientist
Collection and analysis of data pertaining to Nutrition.
Technology Stack and Algorithms :
Technology Stack and Algorithms used :
Python : Machine Learning – Regression Model
Python Libraries : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow
ML Algorithms Used : Unsupervised Learning Algorithms
Steps Used: Descriptive Analysis and Predictive Analysis for all steps of variable description, exploratory data analysis, data processing and pre-processing, feature selection, model building, and evaluation of models are done using Python and R.
Libraries used: Pandas, Scikit-learn, Numpy, Keras, Tensor flow

ii).Sales analysis, customer and retailer analysis :November 2023 –Dec2023
Aims and Objectives :
✓ To understand the sales numbers of a multi-national retail company and to make them grow by analyzing the stock, price, number of items sold and quality rating for various categories and for various products, retailed by various vendors and to find the sales trends.
✓ To gauge and execute suitable sales techniques, to scale up the sales numbers.
✓ To analyze and evaluate the inventory status for various categories, subcategories, and items.
✓ To carry out meticulous exploration of vendor sales and inventory stocks, which will enable the management to formulate schemes to improvise sales in low profit periods and confirm that the inventory has enough stock to meet the customer requirements
  Methodology Adopted – Creation of Various Dashboards:
✓ **Dashboard 1 :Inventory Performance Dashboard.**
✓ **Dashboard 2 :Retailer Analysis Dashboard**

2.Hansa Cequity Pvt Ltd-From March 2022-July 2022

Role : Consultant Data Scientist

Job Profile: Using Python, R, Machine Learning Algorithms . Regression Techniques to Perform Analysis of Mutual Fund Market Research, Mutual fund Investment ,Market Mixed Modelling, Customer Segmentation and Profiling, Customer 360 view, Customer Loyalty Index, Agent Rehiring Model

Technology Stack and Algorithms used: Exploratory Data Analysis, Data Pre-processing, Data Manipulation, Data Visualisation, Linear Regression, Logistic Regression, Decision Tree, Random Forest, RFM Analysis,D Duplication of RE Data Set

Python Libraries used : Numpy, Scikit-learn, pandas, matplotlib, keras, tensor flow,

R , MySQL Work Bench, MS Excel, MySQLite3.0,

3.RGN Price & Co- Chartered Accountants, Mumbai – From Dec 2021 till Feb 2022

Role :Data Analyst - Processing of commercial data related ti shioibg and billing using MS Excel, MS Access,MS Powerpoint

4. Innodatatics, Hyderabad (August 2020 –Dec 2021 Work from Home

Data Science Analyst Intern in Life Science and Health Care Projects
Lead roles in the 4 completed Data Science Projects in Life Science and Health Care as referred above.

----

5 SMAC Consultancy, Mumbai (Jan 2020 to Mar 2020

Roles and Responsibilities - Freelancer Data Analyst
- Worked on Project related to Management studies
- Worked on Hypothesis Testing, p-value analysis, Chi. Square Testing, Auto Report for Data Analysis

----

6. Hydrogreen Agribusiness & Land Cultivators LLP,Mumbai (Apr 2019 - Dec 2019)

Roles and Responsibilities   Data Analyst - Plant / Leaf Analysis
Project Description
Project Domain Agri Business Sector
Detection of Plant Health
Problem definition of detecting Good,Bad and Blight leaves from a file containing approximately 3000 plant images of an agricultural farm.
Method Adopted
In this project, hyperspectral image processing, and analysis techniques was used to detect the plant health which helps in better agricultural produce Since, I was new to the field I took the camera literature from the book 'Hyperspectral imaging technology in food and agriculture' by Series Editor Gustavo V. Barbosa-Ca´novas,
Washington State University, USA.

Skills Used - MATLAB, SCILAB, Python, R, Data Visualisation Microsoft Power point
Libraries Used -MATLAB Imaging, Numpy, Scikit-learn, pandas, matplotlib, keras,

----

7.Cytel Statistical Software & Services Pvt Ltd, Navi Mumbai (June 2018 – Mar 2019)

Project Trainee cum Intern in SAS - Clinical Programming at Cytel's Clinical Professional Laboratory
Worked on CRF Annotations, QCof SDTM Datasets Development, QC of LB Dataset Generations, QC of SDTM Tables

----

8.Central Drug Research Institute, Lucknow (Feb 2017-March 2018)

Project Intern- Online - Work From Home-Computational Biology/ Bioinformatics
Job Responsibilities -
i. Systematically performing in-depth research on the matches between Human Genome Genes as subject and
Genes of other species as query.
ii. Performing Genomics of NF1 gene-Exploring homology incoding and non-coding regions,Overlapping genes,
Adjoining genes,RNA Regulatory Domains, Enhancer and Promoter Regions using R, BLAT, BLAST, Perl.
iii. Systematically conducting computational simulations for Cell Signalling pathways to analyze dynamics of various proteins in network.
Skills Used- R,PERL, Blast
Operating Systems- LINUX, Ubuntu

9.ArrayGen Technologies Pvt Ltd, Pune (May 2016-Oct 2016)
 Project trainee cum Bioinformatics Internship.
 Skills Used -R,PERL,Blast
 Operating System - LINUX Ubuntu
 Roles and Responsibilities
i. Worked on bioinformatics tools such as Vitcomic Package, edgeR, DESeq2, DESeq, Cummerbund, ggplot2 plotting (Client's work), Boxplot, Density Plot, Heatmaps, Volcano Plots, Genomic Studio, Methylumi DataSets, BMIQ,SWAN, methylumiR, TAPIR and psRNATarget
ii. Accurately used FASTQ format like Fasta conversion, BLAST, MIREAP
iii. Meticulously conducted gene expression processes such as WGCNA, CoExpress, endmemo, minfi,
minfidata, DNA-Methylation and gene coexpression..Performed comparative studies such as edgeR,
DESeq2, DESeq Analysis. BSmooth, Bi Seq and normalization within arrays & between arrays.

10.Kotak Education Foundation, Chembur, Mumbai (Aug 2015-Oct 2016)
 Honorary Visiting Faculty-Applied Maths for Engineering. Students

----
Training Certifications in Data Science

1 E&ICT Academy, IIT-Kanpur Certification course on Advanced Certification in AI & ML-Oct 2023 .**(Oct 2022- Sep 2023)**
2 AI Tools Mastery Program and Effective Prompt writing in Chat GPT conducted by be10X-Oct 2023.
3 Chat GPT for Excel in Sept 2023 issued by Udemy Academy-Oct 2023
4 Certified Deep Learning Expert issued by Edvancer Eduventures in OCT 2023
5 Certificate of Excellence in Data Analysis using SQL in Sept 2023 issued by Edvancer Eduventures.
6 Certified Tableau Professional in July 2023 issued by Edvancer Eduventures
7 Advanced Python for Data Science – Almabetter July 2022
8 R for Data Science – IBM Certified- Intellipaat – Oct 2021
9 Data Science Master's Architect Program IBM Certified– Intellipaat – Jan 2021- May 2021
10.Natural Language Processing with Python – Udemy Academy-Nov 2020-Dec 2020
11 Data Science course in Life Science & Health Care analytics- 360DigitMg- July 2020-Sep2020
12 Software Construction in Java – MIT edX- Sep 2016- Dec 2016
13 Clinical Statistical Programming Training and Project Readiness–June 2018-Mar 2019
14Advanced Bioinformatics – Microarrays & NGS – Arraygen Technologies- 2016
15 Object Oriented Programming - C++ NIIT Apr 2016-Sep 2016

AWARDS AND RECONITION

- Secured 10 Bronze Medals in in a row in the ongoing Data Science Kaggle Competitions in Sept 23- Feb 2024
-
- Won multiple Prizes and accolades from Kaggle, for publishing solutions to complex problems and several in the form of views and forks for my notebook.
- Semifinalist at TechGig Coding Competition 2021
- Secured All India Rank 41 in GATE Exam in 2009◇Secured 305/340 in GRE in 2011.
- Secured INRIA International Internship Award, INRIA, Paris, France- MTech 2011-2012
- PUBLICATIONS
- Publication in PLOS One Journal :
- Worked on Research paper publication-in Computational Biology (2011-13) Topic -
- Noise Propagation in Two- Step Series MAPK Cascade

- Link : https://doi.org/10.1371/journal.pone.0035958

Personal DETAILS

_ **Linkedin Profile** - https://www.linkedin.com/in/gopalakrishnan-kumar-a73301110/
_ **Github Profile: https://github.com/Gopalakrishnan-Kumar/Python-for-Data-Science**
_ **Kaggle Profile: https://www.kaggle.com/gopalkk1**
-   **Kaggle Notebook Expert seen in Profile**
_ **Date of Birth -** 29th July, 1986
_ **Nationality** – Indian
_ **Father's Name:** Kumar Krishnan
_ **Alternate Mobile No :**9819043881
_ **Marital Status**: Single
_ **Languages Known** - English, Hindi, Marathi,Tamil
_ **Hobbies** - Reading articles on Data Science, listening to Music, playing violin