Contact: +918967821155

Permanent Address:

Vivekananda Para, Bairatiguri, Ward No. 6, Dhupguri, Jalpaiguri, West Bengal, India – 735210

E-mail ID:

Kaushik.das220797@gmail.com

LinkedIn ID:

https://www.linkedin.com/in/kaushik-das-0736831b2

HackerRank ID:

https://www.hackerrank.com/profile/kaushik_das_exp1

Top Skills:

Programming:

1. *Python*
2. *R-Programming*
3. *SQL*

Machine Learning

Deep Learning

Statistical Skills

1. *Probability*
2. *Hypothesis Testing*
3. *Quality Control*

Data Wrangling

Data Pre-processing

Problem Solving

Research

Story Telling

Data Visualization

BI Tools:

1. *Power BI*
2. *Tableau*

Data Analysis

Interpersonal Skills

Mathematical Concepts

1. *Algebra*
2. *Calculus*
3. *Optimization Techniques*

Cloud Computing

Proficient in English

Kaushik Das

My aspiration is to excel as a great Data Scientist, driven by a profound enthusiasm for working with data. With a robust understanding and implementing various Machine Learning (ML) and Deep Learning (DL) algorithms, along with a solid grasp of the statistical and mathematical principles underpinning them. Proficient in feature engineering and data preprocessing, I also command a strong foundation in statistical concepts such as quality control and hypothesis testing. With over a decade of coding experience across multiple programming languages including Python and R, I am well-equipped to tackle diverse challenges in AI ML projects and research and development. Coupled with effective communication skills and fluency in English, I am keen to contribute to cutting-edge initiatives in the field. Driven by a relentless pursuit of improvement, I am a dedicated and a life-long learner committed to continual growth.

Internships:

1. Data Analyst at Pahal Financial Services pvt. ltd. (25th Sep.2023 – 23rd Dec. 2023)
2. Business Analyst at Bajaj Allianz Life Insurance (Dec.2021-Apr.2022)
3. Student Associate at CHRIST Consulting (Aug. 2020 – Oct. 2021)

Educational Qualifications:

Degree	Specialization	Institution	Passing Year	%age
M.Sc.	Data Science	Christ University	Ongoing	60%
B.Sc.	Computer Science, Statistics & Mathematics	Kristu Jayanti College Autonomous	2020	70%

Certifications:

1. Professional Certificate Course in AI and Machine Learning from E & ICT Academy, IIT Kanpur (A joint initiative of MeitY & IIT Kanpur)
2. Earned 5th Star in “Python” on HackerRank
3. Earned 5th Star in “Problem Solving” on HackerRank
4. Cleared the assessment for the skill “Problem Solving (Basic)” and in “Python (Basic)” on HackerRank
5. Completed training program on “Applied Data Science with Python” – Simplilearn
6. Completed training program on “Machine Learning” – Simplilearn
7. Earned LinkedIn Skill Assessment Badges in “Power BI”, “MS Excel” & “MS Power Point”
8. Tableau A to Z & Power BI – Udemy
9. Inferential Statistics – Coursera
10. AWS Foundation Course

Projects:

1. House Loan Data Analysis (ML and DL concepts)

- Checked for missing values for each feature and removed the features and records whose null value percentage was more than 50% and removed duplicate records. Performed basic statistical calculations on numeric features.
- Identified categorical and ordinal features, conducted one-hot encoding and ordinal encoding respectively and removed all the nominal features.
- Balanced the dataset by applying SMOTE technique
- Trained the model with classification algorithms such as Logistic Regression, Decision Tree Classifier, Random Forest Classifier (Ensembled technique), Gaussian NB Classifier, K-NN Classifier and MLP Classifier and found the best model by comparing Accuracy scores, Sensitivity scores (for reduction of False Positive values) and AUC-ROC score for every model and obtained the best model.
- Performed Hyperparameter tuning with best performing classifier for different parameters of hidden layer sizes, activation functions, solvers and alpha values using GridSearchCV and selected the best combination of hyperparameters that further increases the Sensitivity score.
- Libraries used: pandas, imblearn, sklearn, seaborn.

2. Creating Cohorts of Songs (Spotify Music Recommendation System)

- Used dataset from Spotify's API about all albums for the Rolling Stones
- Inspected, cleaned, performed EDA and feature engineering on the dataset and discovered how the popularity of songs changed on various features.
- Handled imbalanced data using SMOTE technique
- Used Variance Inflation Technique (VIF) to remove multicollinearity of the data
- Used PCA (Principal Component Analysis) for dimensionality reduction of the data such that the complexity of the dataset is reduced (by reducing the number of features) while keeping the most important properties of the original data.
- Identified right number of clusters using Silhouette Score and Davies Bouldin Score
- Created cohorts of different songs that would aid in the recommendation of songs to users based on popularity using K-means clustering.
- Programming Language used: Python
- Libraries used: pandas, NumPy, plotly, seaborn, matplotlib, sklearn, imblearn

Hobbies and Interests:

Meditation

Science and Technology

Geopolitics

Personal Details:

DOB: 22nd July, 1997

3. Automating Port Operations (using CNN)

- Using Deep Learning techniques to build an automatic reporting system that recognizes the various kinds of boat images and classifies them.
- Model 1: Built CNN (Convolutional Neural Network) to classify the boat using Keras with two filters, one pooling layer and three Dense layers. Used relu and softmax as activation functions
- Compiled the model with Adam optimizer and categorical cross entropy loss (multi-class classification) with metrics accuracy, precision and recall.
- Trained the model with 20 epochs, evaluated the model
- Plotted heatmap of the confusion matrix and generated classification report.
- Model 2: Built a lightweight model MobileNetV2 using Keras with the aim of deploying the solution on a mobile device using transfer learning
- Used CNN network with Keras and used MobileNetV2 model as the first layer, added pooling, dropout, Batch Normalization layers, activation function used relu and softmax in the output layer due to multi-class classification.
- Compiled the model with Adam optimizer and used categorical cross entropy loss and metrics accuracy, precision and recall
- Trained the model with 50 epochs and Early stopping while monitoring validation loss.
- Evaluated model on test images and obtained test loss and accuracy and plotted Train loss Vs Validation loss and Train accuracy Vs Validation accuracy.
- Compared the results of both the models and made a detailed report mentioning the outcomes
- Libraries used: TensorFlow, matplotlib, cv2, os, NumPy, glob, seaborn

4. Lending Club Loan Data Analysis (DL concept)

- Checked for missing values and did a statistical check for every features for different central tendencies, measures of skewness and dispersion
- Performed feature transformation such as replacing missing value of the features with median and transformed categorical values into numerical values (discrete)
- Performed Exploratory data analysis of different factors of the dataset to understand the spread of data for different factors and to check if there are any outliers present.
- Checked for imbalanced data and handled imbalanced data using SMOTE technique.
- Checked the correlation between features and dropped those features which have a strong correlation to reduce multicollinearity.
- Performed PCA (Principal Component Analysis) for dimensionality reduction and reduced the dimensions to five.
- Used MLP (Multi-layer Perceptron) algorithm to train the model
- Chose the best hyperparameters from a pool of different hidden layer sizes, activation functions, alpha values to obtain a model with high accuracy and high precision (in order to reduce false positive cases i.e. to minimise the number of potential defaulters that were provided a loan)
- Evaluated the model and obtained train accuracy of 92.1%, test accuracy of 91.7% and precision of 94.3%
- Libraries used: pandas, sklearn, seaborn, matplotlib, plotly, imblearn, hyperopt

5. Employee Turnover Analytics

- Used HR data from Portobello company from Kaggle to predict retention of employees using K-Means and Cross-Validation techniques.
- Performed data quality checks, data pre-processing and EDA (Exploratory Data Analysis) and understood what factors contributed most to employee turnover
- Performed clustering of Employees who left the organisation using K-Means clustering with three clusters.
- Handled imbalanced data using SMOTE technique
- Performed 5-fold Cross Validation model training using Logistic Regression Model, Random Forest Classifier Model and Gradient Boosting Model and identified the best model by comparing various evaluation metrics using Recall value (to minimise False Negative values in confusion matrix)
- Suggested various retention strategies for targeted employees by using the best model, then predicted the probability of employee turnover in the test data and categorized the data into four zones.
- Programming language used: Python
- Libraries used: pandas, sklearn, seaborn, matplotlib, plotly, yellowbrick, imblearn.

6. Analysis on Non-renewed Customers Data

- Mapping of customer's data for last six months that had opted for Joint Liability Loan
- Working on creating datasets and dashboards for non-renewed customers to find out how likely their policies can be renewed. Pointing out actionable insights.
- Making Daily Member Login Report, Collection vs Disbursement Report and Target vs Achievement Report for JLG customers.
- Tools used: Python, MS Excel and Power BI

7. Cybersecurity: Machine Learning model

- Working on research for detection of cyber-attacks using supervised Machine Learning Algorithms
- Tools used: Python

8. Quality Portfolio Analysis using Dashboards

- Find out growth/fall of the issued insurance policies in a given period
- Monitor performances of policies sourced in different categories
- Highlight the Actionable insights based on the analysis done
- Tools used: R, SQL, MS Excel and Tableau

9. Performing Hypothesis testing of Reaction time

- Took samples of reaction time using a game for both boys and girls from the population of college
- Conducted hypothesis testing to find if there is any significant difference between the reaction time of boys and girls or not using z-test.
- Tools used: MS Excel, Statistical techniques.

10. College Study Project

- Conducted surveys on college going students and professors on Virtual learning in higher education (globally)
- Worked on Qualitative Research and Quantitative Data Analysis
- Tools used: MS Excel and Python

Achievements:

1. 1st prize in Intra-Collegiate Science Exhibition in Statistics (Galaxia 2018), Kristu Jayanti College, Bangalore (in 2018)

2. 2nd Prize in Inter-Collegiate Statistics fest (Frequentia 2018), Treasure Hunt Event, in St. Joseph College, Bangalore (in 2018)

3. 2nd Prize in Intra-Collegiate Science Exhibition in Mathematics category (Galaxia 2019), Kristu Jayanti College, Bangalore (in 2019)

4. 2nd Prize in Inter-collegiate Statistics fest in Auction Event (Inference 2019), CHRIST (Deemed to be University), Central Campus, Bangalore (in 2019)

5. 1st Prize in Intra-collegiate Data Science (Online) fest (Technophilia 2020) in CHRIST in Analytica Event (Deemed to be University), Lavasa Campus, Pune (in 2020)

6. 1st Prize in Inter-Collegiate Statistics (Online) fest (Estimato 2020) in Statistical Paradox Event, Kristu Jayanti College, Bangalore (in 2020)

7. 2nd Prize in Inter-collegiate Data Analysis Fest (Online) (Data Express) in Mount Carmel College Autonomous, Bengaluru (in 2021)