

Pankaj Kumar Yadav

Dedicated data science enthusiast with a strong desire to use data to discover and solve business problems. Proficient in predictive modeling, data processing, data mining, problem-solving thinking and creativity in identifying solutions.

✉ dev.pankajkumaryadav@gmail.com

☎ 8538992331

🌐 linkedin.com/in/pankaj-kr-yadav

🐙 github.com/yadavkumarpankaj

WORK EXPERIENCE

Data Science Trainee

AlmaBetter School

05/2022 - Present

Bangalore, India

Key Responsibilities and accountabilities:

- Gained in-depth understanding of skills such as **Python**, **SQL**, and **Machine Learning** and improved **story-telling** skills by delivering presentations or **visualizations** regarding insights generated while working on projects
- Efficiently used tools such as **Google Colab**, **Tableau**, and **PowerBI** to implement real world projects such as **EDA on World Bank Global Education Analysis**, Yes Bank Stock Closing Price Prediction etc. Worked on SQL case studies like Danny's Diner etc.
- Achieved **Star Student Badge** Award given to **top four** performers for working on individual project and demonstrating work through **skillful presentations**.

PROJECTS

Online Retail Customer Segmentation

AlmaBetter Verified Project 

11/2022 - 12/2022

Tags: *Clustering, K-Means, Hierarchical, RFM Score, Elbow Method, Silhouette Score, Pandas*

- Developed **clustering model** to identify **customer segments** on retail transactional dataset to **tailor marketing efforts**.
- **Explored** data and created feature called **RFM Score** to know **likelihood** of customers purchasing again.
- Implemented **K-Means** to cluster data into **3 segments** which was validated using **elbow method** and **silhouette score of 0.3** showing samples in one cluster are close to their neighboring points from same cluster and far from other cluster.
- **K-Means** with **3 clusters** helped improve marketing efforts & generate more sales.

Cardiovascular CHD Risk Prediction

AlmaBetter Verified Project 

10/2022 - 11/2022

Tags : *Logistic Regression, Classification, Hyperparameter Tuning, SVM, Random Forest.*

- Developed **classification model** to predict **ten-year risk of coronary heart disease** for different age groups to **build preventive methods** against CHD.
- Performed **univariate**, **bivariate** and **multivariate analysis** for detecting **outliers** and **missing values**, combined with **feature engineering** to draw out best set of inputs and fixed class imbalance via **SMOTE**.
- Performed **random forest**, **support vector machine**, **xgboost**, and **logistic regression** with **cross validation(CV)** algorithms with different **evaluation metrics** and applied **hyperparameter** tuning to them & predicted CHD risk with **F1 score of 91 %**.
- **Causes** that are most likely to **result** in **CHD** are easily determined by doctors, and **prevention** of **10-year CHD** can be suggested to **patients** and public.

Yes Bank Stock Closing Price Prediction

AlmaBetter Verified Project 

08/2022 - 09/2022

Tags: *matplotlib, seaborn, linear regression, Lasso Regression, feature engineering, XGBoost, EDA*

- Built **regression model** to accurately predict **stock market** returns that is **highly volatile** and **non-linear** in nature which lead to financial gain.
- Performed **Exploratory data analysis**, Imputed missing values, encoded categorical features, handled outliers, and checked **multicollinearity** combined with **Feature engineering** to identify optimal inputs.
- With introduction of **machine learning**, **time series forecasting**, and increased **computational capabilities**, programmed **methods of prediction** have proved to be more **efficient in predicting** stock prices
- In this work we use **linear regression** technique, **lasso regression**, **ridge regression**, **elastic net regression**, and **XGBoost Regression** technique. these five models have used.

TECHNICAL SKILLS

Language, Database & ML Libraries

Python, SQL, Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, SciPy, TensorFlow.

Frameworks and Platforms

Jupyter Notebook, Google Colab, Postgre SQL, MySQL, MS Word, MS Excel, Tableau, PowerBI, Github.


Algorithms

Linear Models, Decision tree, SVM, Ensemble Models, KNN, Naive Bayes, Clustering Algorithms, Recommendation System.


Professional Skills

Data Science, Data Engineering, Data Analytics, Business Analytics, ML Deployment


COURSES

Full Stack Data Science (05/2022 - Present) 

AlmaBetter School, Bangalore


Machine Learning (07/2022 - 10/2022) 


Deeplearning.AI, Coursera

IBM Data Analyst (04/2022 - 09/2022) 

IBM, Coursera

ACHIEVEMENTS

All Earned Badges From IBM and Coursera 

Earned Python and SQL certifications and gold badges from Hackerrank. 

PUBLICATIONS

Medium

Medium Blogs on DataScience 

2022

EDUCATION

B.com

Indra Gandhi National Open University

06/2017 - 06/2021

60%

XII- Higher Secondary

R.N Singh Senior Secondary Public School

05/2015 - 06/2017

63%

INTERESTS

Workout

Gaming

Meditation