

# Data Science & Machine Learning Notebook Collection (English + বাংলা)

---

For my own learning, in my own words, in my own way! | নিজের শেখার জন্য, নিজের ভাষায়, নিজের মতো করে!

---

## Table of Contents | সূচিপত্র

- Quick Start | দ্রুত শুরু
  - How to Use | কিভাবে ব্যবহার করবেন
  - Expanded Notes by Folder | ফোল্ডারভিত্তিক বিস্তারিত নোট
  - Learning Tips | শেখার টিপস
  - FAQ & Resources | সাধারণ প্রশ্ন ও রিসোর্স
- 

## Quick Start | দ্রুত শুরু

- Python 3.7+
- Jupyter Lab/Notebook
- Libraries: pandas, matplotlib, seaborn, scikit-learn, pandas-profiling, gspread
- পাইথন ৩.৭+, জুপিটার ল্যাব/নোটবুক, pandas, matplotlib, seaborn, scikit-learn, pandas-profiling, gspread

### Install | ইনস্টল করুন:

```
git clone https://github.com/your-username/your-repo-name.git
cd your-repo-name
pip install pandas jupyterlab matplotlib seaborn scikit-learn pandas-profiling
gspread
jupyter lab
```

## How to Use | কিভাবে ব্যবহার করবেন

- Browse folders by topic (e.g., Pandas, Linear Regression, etc.) | টপিক অনুযায়ী ফোল্ডার দেখুন
  - Open any `.ipynb` notebook in Jupyter | যেকোনো `.ipynb` নোটবুক খুলুন
  - Each notebook has step-by-step code, explanation, and real datasets | প্রতিটি নোটবুকে ধাপে ধাপে কোড, ব্যাখ্যা ও ডেটাসেট আছে
- 

## Expanded Notes by Folder | ফোল্ডারভিত্তিক বিস্তারিত নোট

01\_pandas — Data Handling with Pandas

---

## What is Pandas? | Pandas কী?

- English: Pandas is a powerful Python library for data manipulation and analysis.
- বাংলা: Pandas হলো পাইথনের শক্তিশালী ডেটা ম্যানিপুলেশন ও অ্যানালাইসিস লাইব্রেরি।

### Key Concepts | মূল বিষয়:

- Creating DataFrames from lists, dicts, or CSV/Excel files
- Subsetting, sorting, and describing data
- Data profiling for quick insights
- Importing data from Google Sheets

### Example | উদাহরণ:

```
import pandas as pd
df = pd.read_csv('Screen Time Data.csv') # Load CSV | CSV লোড করুন
print(df.head()) # Show first rows | প্রথম কয়েকটি সারি দেখুন
```

### Tips | টিপস:

- Use `.info()`, `.describe()`, `.shape` to understand your data | ডেটা বোঝার জন্য এগুলো ব্যবহার করুন

---

## 02\_types of var — Types of Variables in Data Science

### Understanding Data: Key Dimensions | ডেটা বোঝার মূল দিক

- **Data Types | ডেটার ধরন:**
  - Numerical (Discrete, Continuous) | সংখ্যাগত (ডিসক্রিট, কন্টিনিউয়াস)
  - Categorical (Nominal, Ordinal, Binary) | ক্যাটাগরিক্যাল (নমিনাল, অর্ডিনাল, বাইনারি)
  - Text, Date/Time | টেক্সট, তারিখ/সময়
- **Measurement Scale | মাপার স্কেল:**
  - Nominal, Ordinal, Interval, Ratio | নমিনাল, অর্ডিনাল, ইন্টারভ্যাল, রেশিও
- **Role in Analysis | বিশ্লেষণে ভূমিকা:**
  - Independent (Input), Dependent (Output) | ইনপুট, আউটপুট
- **Distribution | ডিস্ট্রিবিউশন:**
  - Normal, Skewed, Uniform, Bimodal | নরমাল, স্কিউড, ইউনিফর্ম, বাইমোডাল

### Why it matters? | কেন গুরুত্বপূর্ণ?

- Choosing the right statistical methods and visualizations depends on variable types.
- সঠিক বিশ্লেষণ ও ভিজুয়ালাইজেশনের জন্য ধরন জানা জরুরি।

---

## 03\_linear regression — Linear Regression & Related Concepts

### Salary Prediction using Linear Regression

- English: Predict salary from years of experience using linear regression. Steps: Data load, EDA, model building, evaluation, and result analysis.
- বাংলা: লিনিয়ার রিগ্রেশন দিয়ে অভিজ্ঞতার বছর থেকে বেতন অনুমান। ধাপ: ডেটা লোড, EDA, মডেল তৈরি, মূল্যায়ন, ফলাফল বিশ্লেষণ।

### EDA Example | EDA উদাহরণ:

```
import pandas as pd
df = pd.read_csv("Salary Data.csv")
print(df.head())
```

### Linear Regression Model | লিনিয়ার রিগ্রেশন মডেল:

```
from sklearn.linear_model import LinearRegression
reg = LinearRegression()
reg.fit(X, y)
```

### Loss & Cost Functions

- **MSE (Mean Squared Error):** Measures average squared difference between actual and predicted values.
- **MAE (Mean Absolute Error):** Measures average absolute difference.

### Formula | সূত্র:

```
[
MSE = \frac{1}{N} \sum (y - \hat{y})^2
]
[
MAE = \frac{1}{N} \sum |y - \hat{y}|
]
```

### R-squared & Model Evaluation

- **R-squared:** Proportion of variance in target explained by the model.
- **Formula:**

```
[
R^2 = 1 - \frac{SS^{*}\{res\}}{SS^{*}\{tot\}}
]
```

## 04\_Feature Engineering — Feature Engineering

### What is Feature Engineering? | ফিচার ইঞ্জিনিয়ারিং কী?

- Transforming raw data into features for ML (encoding, scaling).

- কাঁচা ডেটাকে মডেলিংয়ের জন্য ফিচারে রূপান্তর করা (এনকোডিং, স্কেলিং)।

### Key Techniques | মূল টেকনিক:

- Label Encoding, One-Hot Encoding, Binary Encoding, Ordinal Encoding
- Min-Max Normalization, Standardization (Z-Score)

### Example | উদাহরণ:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['encoded'] = le.fit_transform(df['category'])
```

---

## 05\_project\_multiple\_linear\_regression — Multiple Linear Regression Project

### Project Overview | প্রজেক্ট ওভারভিউ:

- Predicting profit using multiple features (Marketing Spend, Administration, Transport, Area).
- Data cleaning, one-hot encoding, train-test split, model fitting, scoring.

### Note:

- Always check for nulls or zeros in data before modeling.
- মডেল করার আগে ডেটা null বা 0 কিনা চেক করুন।

---

## 06\_Gradient Descent — Gradient Descent

### What is Gradient Descent? | গ্র্যাডিয়েন্ট ডিসেন্ট কী?

- Optimization algorithm to minimize cost functions.
- কস্ট ফাংশন মিনিমাইজ করার অপটিমাইজেশন অ্যালগরিদম।

### Steps | ধাপসমূহ:

1. Initialize parameters (e.g.,  $m, c = 0$ )
2. Update using gradients
3. Repeat for set iterations

### Example | উদাহরণ:

```
m, c = 0, 0
learning_rate = 0.001
for i in range(1000):
    y_pred = m*x + c
    md = -(2/n) * sum(x*(y-y_pred))
    cd = -(2/n) * sum(y-y_pred)
```

```
m = m - learning_rate * md
c = c - learning_rate * cd
```

---

## 07\_Decision Tree — Decision Tree

### What is a Decision Tree? | ডিসিশন ট্রি কী?

- Splits data into branches to make decisions.
- ডেটাকে ভাগ করে সিদ্ধান্ত নেয়।

### Key Concepts | মূল বিষয়:

- Information gain, entropy, tree pruning.

---

## 08\_Confusion Matrix — Confusion Matrix

### What is a Confusion Matrix? | কনফিউশন ম্যাট্রিক্স কী?

- Table to evaluate classification model performance.
- ক্লাসিফিকেশন মডেলের পারফরম্যান্স যাচাইয়ের টেবিল।

### Metrics | মেট্রিক্স:

- Accuracy, Precision, Recall, F1-score

---

## 09\_K-Fold Cross Validation — K-Fold Cross Validation

### What is K-Fold CV? | K-Fold ক্রস ভ্যালিডেশন কী?

- Splits data into k parts to validate model robustness.
- ডেটাকে k ভাগে ভাগ করে মডেল যাচাই করা।

---

## 10\_Logistic Regression Classification — Logistic Regression

### What is Logistic Regression? | লজিস্টিক রিগ্রেশন কী?

- Used for binary classification (yes/no, 0/1).
- দুই ক্লাসের (হ্যাঁ/না, ০/১) প্রেডিকশনের জন্য ব্যবহৃত।

---

## 11\_Support Vector Machine — SVM

### What is SVM? | SVM কী?

- Finds the best boundary (hyperplane) to separate classes.
- ক্লাস আলাদা করতে সবচেয়ে ভালো বাউন্ডারি বের করে।

## 12\_Predicting Heart Disease using Random Forest — Random Forest

### What is Random Forest? | র্যান্ডম ফরেস্ট কী?

- Ensemble of decision trees for better accuracy.
  - অনেক ডিসিশন ট্রির সমষ্টি।
- 

## 13\_Save ML Model Pickle vs Joblib — Model Saving

### What is Model Saving? | মডেল সংরক্ষণ কী?

- Save and load ML models for reuse.
  - মডেল সংরক্ষণ ও পুনরায় ব্যবহার।
- 

## 14\_K-Nearest Neighbors Algorithm in ML — KNN

### What is KNN? | KNN কী?

- Predicts by looking at the k closest data points.
  - কাছে k ডেটা পয়েন্ট দেখে প্রেডিক্ট করে।
- 

## 15\_ What is CountVectorizer in Python & How CountVectorizer Work — CountVectorizer

### What is CountVectorizer? | CountVectorizer কী?

- Converts text to numeric features for ML.
  - টেক্সটকে নিউমেরিক ফিচারে রূপান্তর।
- 

## 16\_Naive Bayes Algorithm with Python — Naive Bayes

### What is Naive Bayes? | ন্যেভ বেইজ কী?

- Probabilistic classifier based on Bayes' theorem.
  - Bayes' theorem ভিত্তিক সম্ভাব্যতাভিত্তিক ক্লাসিফায়ার।
- 

## 17NLP Project Spam eMail Detection with Naive Bayes Classifiers — NLP Spam Detection

### What is NLP Spam Detection? | NLP স্প্যাম ডিটেকশন কী?

- Detects spam emails using NLP and Naive Bayes.
- NLP ও Naive Bayes দিয়ে স্প্যাম ইমেইল শনাক্তকরণ।

### Key Steps | ধাপসমূহ:

- Data cleaning, feature extraction (CountVectorizer), model training, evaluation.
-

## Learning Tips | শেখার টিপস

- Practice by running code | কোড চালিয়ে প্র্যাকটিস করুন
  - Visualize data and results | ডেটা ও রেজাল্ট ভিজুয়ালাইজ করুন
  - Read documentation | ডকুমেন্টেশন পড়ুন
  - Take notes | নোট নিন
- 

## FAQ & Resources | সাধারণ প্রশ্ন ও রিসোর্স

- **ModuleNotFoundError?** Install all libraries: `pip install ...` | সব লাইব্রেরি ইনস্টল করুন
- **Jupyter won't start?** Try `jupyter notebook` | চালু না হলে `jupyter notebook` ব্যবহার করুন
- **Data file not found?** Run from project root | প্রজেক্ট রুট থেকে চালান
- **Google Sheets import?** See [Google Sheets notebook](#)

### Resources | রিসোর্স:

- [Pandas Docs](#)
  - [Scikit-learn Docs](#)
  - [Matplotlib Docs](#)
  - [Kaggle Datasets](#)
  - [Google Colab](#)
- 

Happy Learning! | শুভকামনা! 🍀