Assignment: Data Science Application Assignment
Pakistan's Largest E-commerce Dataset

Course: Data Science Applications Instructor: Zuhaib Hussain Butt

Objective

The objective of this assignment is to assess students' knowledge and practical skills in applying the complete data science workflow on a real-world e-commerce dataset. Students must demonstrate understanding of data preprocessing, feature engineering, predictive modeling, evaluation, business interpretation, and application deployment considerations.

Dataset

- Pakistan's Largest E-commerce Dataset (Kaggle)
- The dataset includes product details, customer demographics, categories, order information, price, reviews, and transaction metadata.

Submission Format

- Final report in **PDF** format only.
- Must include headings listed below.
- Code is **not** required inside the report, but methodology must be clearly described.
- Use any programming language or data science tool of your choice.

Report Headings (Structure)

Students must structure the final PDF report using the following section headings:

1. **Introduction**

   1.1. Overview of dataset

1.2. Problem statement(s) defined by the student

1.3. Relevance to e-commerce industry

2. Data Understanding

2.1. Description of fields

2.2. Type of variables (numerical, categorical, text)

2.3. Initial observations

3. Data Cleaning & Preprocessing

3.1. Handling missing values

3.2. Outlier detection (describe approach)

3.3. Data type correction

3.4. Normalization/standardization (if needed)

3.5. Data integration and filtering steps

4. Feature Engineering

4.1. New features created

4.2. Encoding of categorical features

4.3. Text processing steps (if applied)

4.4. Feature selection or dimensionality reduction (optional)

5. Predictive Modeling Students must choose and justify **one** predictive task from:

- Price prediction (regression)
- Product-category classification
- Delivery-time prediction
- Customer segmentation (unsupervised)
- Fraud detection (classification)

For the chosen task, include:

5.1. Choice of model(s) and justification

5.2. Model training process

5.3. Hyperparameter tuning strategy

6. Model Evaluation

6.1. Train-test split summary

6.2. Evaluation metrics (appropriate to task)

6.3. Comparison of models (if multiple tested)

6.4. Interpretation of results

7. **Business Interpretation** Explain:
   - What your model's results mean for the business
   - How this predictive system can improve operations
   - How accuracy or prediction insights translate to revenue or efficiency
8. **Application Deployment Considerations** Discuss:
   - Model serving options (API, web app, batch pipeline)
   - Required system architecture
   - Real-world constraints (latency, cost, scalability)
   - Ethical considerations (data bias, fairness)
9. **Conclusion**

9.1. Summary of findings

9.2. Final recommendation for e-commerce business

9.3. Limitations and future improvements

Rubric (Total 50 Marks)

- **Introduction & Problem Definition** – 5 Marks
- **Data Understanding** – 5 Marks
- **Cleaning & Preprocessing** – 8 Marks
- **Feature Engineering** – 7 Marks
- **Predictive Modeling** – 10 Marks
- **Model Evaluation** – 8 Marks
- **Business Interpretation & Deployment Strategy** – 5 Marks
- **Conclusion & Presentation Quality** – 2 Marks

Academic Integrity

All submitted work must be original. Plagiarism or duplicated reports will result in zero marks.