**Problem Statement**

Design, implement and evaluate an end-to-end unsupervised learning model using Keras and TensorFlow. The goal of the project is to develop a system that can identify patterns of anomalies in a given dataset and provide actionable recommendations based on the detected anomalies.

**Dataset description**

TensorFlow Datasets provides a variety of datasets that you can use for anomaly detection tasks. Use the tfds.list\_builders() function to see a list of available datasets, and search for Anomaly Detection Datasets using “search\_results = tfds.list\_builders(search\_text='anomaly')”. Once you identify a dataset suitable for your anomaly detection task, you can load it using the tfds.load() function

**Tasks to be completed**:

**Marks: 50**

**1.** **Data Preprocessing and Exploration (4 marks)**

* Load the chosen dataset and perform initial data exploration to understand its structure and characteristics.
* Preprocess the data, handling missing values, scaling, and encoding categorical features if necessary.
* Split the dataset into a training set and a test set.

**2. Unsupervised Learning Model Development (6 marks)**

* Implement an autoencoder architecture using Keras to learn the normal patterns within the data.
* Train the autoencoder on the training data with the objective of minimizing reconstruction error.
* Use the trained autoencoder to reconstruct the data and calculate the reconstruction error for each instance.

1. **Anomaly Detection (6 marks)**

* Define a threshold for the reconstruction error above which instances are considered anomalous.
* Identify instances in the test set that have a reconstruction error exceeding the threshold.
* Flag these instances as anomalies.

**4.** **Actionable Recommendations (4 marks)**

* For each detected anomaly, analyze the features that contributed the most to the anomaly.
* Based on the specific features and context, provide actionable recommendations or suggestions for further investigation.
* These recommendations could involve suggesting immediate actions or flagging the anomalies for further analysis by domain experts.

**5. Evaluation and Visualization (6 marks)**

* Evaluate the performance of the anomaly detection model using metrics such as precision, recall, F1-score, and ROC curve.
* Visualize the detected anomalies along with the actionable recommendations in a meaningful way.
* Use visualizations and summary statistics to convey insights gained from the analysis.

**6. Model Refinement and Optimization (4 marks)**

* Experiment with different autoencoder architectures, hyperparameters, and reconstruction error thresholds to fine-tune the model's performance.
* Apply techniques such as learning rate scheduling, regularization, and dropout to improve the model's generalization and robustness.

**7. Documentation (10 marks) and 15 mins Presentation (10 marks)**

Document the entire project, including the dataset description, preprocessing steps, model architecture, hyperparameter settings, and evaluation results.

Create a presentation summarizing the project's objectives, methods, findings, and recommendations.