**ML Project Report – Urdu Comment Classification**

**Project Summary:**

We are using Hugging Face Transformers library to create a machine learning model that tells if an Urdu comment is sensitive (1) or not sensitive (0).

**Library Used: Transformers**

* Transformers is a Python library made by Hugging Face.  
  It gives us pre-trained models (like XLM-RoBERTa) that understand many languages.  
  We used this library to train our model on Urdu text.
* These models can understand and work with natural language, like English, Urdu, etc.
* Pre-trained means the model has already learned a lot from reading huge amounts of text, so we don’t need to train from scratch.

**Main Functions Used:**

* AutoTokenizer.from\_pretrained()

Changes text into tokens (Numeric values) so the model can understand it.

* AutoModelForSequenceClassification.from\_pretrained()

Loads a ready-made model for classification (0 or 1).

* Dataset.from\_pandas()

Converts our Excel data into Hugging Face dataset format.

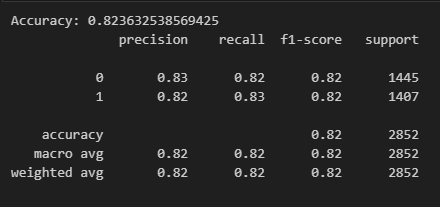
* Trainer

Used to train and evaluate the model.

* TrainingArguments

Settings for training (like number of times the model should learn).

**ScreenShots:**

****

**A black screen with white numbers

AI-generated content may be incorrect.**

**A blue squares with white text

AI-generated content may be incorrect.**