Bank-BERT 

Mrinoy Banerjee

23 April, 2024

AGENDA

- Motivation
- Existing Alternatives
- Problem
- Data
- Models
- Evaluation
- Pipeline
- Demo
- Conclusions
- Future Growth

MOTIVATION

Automated, accurate and personalized analysis of your finances without sharing your financial data.

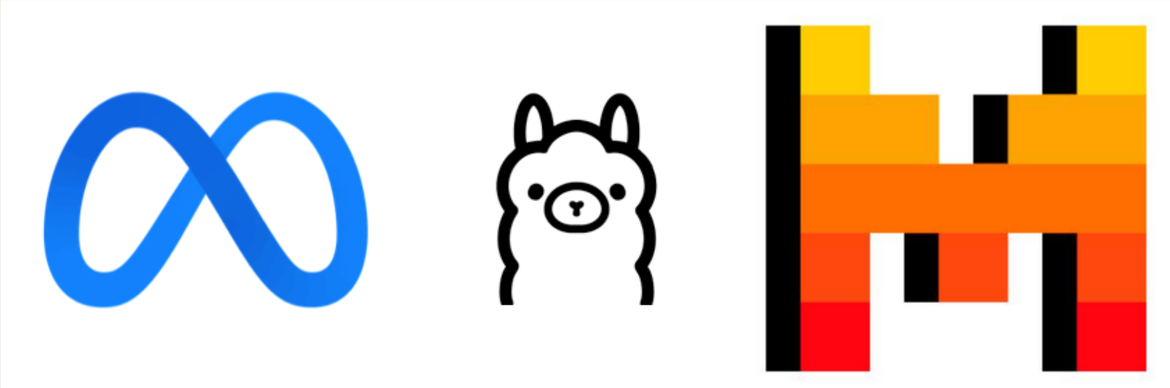
Own your finances!

EXISTING ALTERNATIVES

MERCHANT CATEGORY CODES

- BANKS AND PERSONAL FINANCE APPS HAVE A CLASSIFICATION ACCURACY OF 100%.
- EVERY BUSINESS HAS A MERCHANT CATEGORY CODE.
- EXCEL SPREADSHEETS ALTHOUGH SECURE CAN BE TIME-CONSUMING.

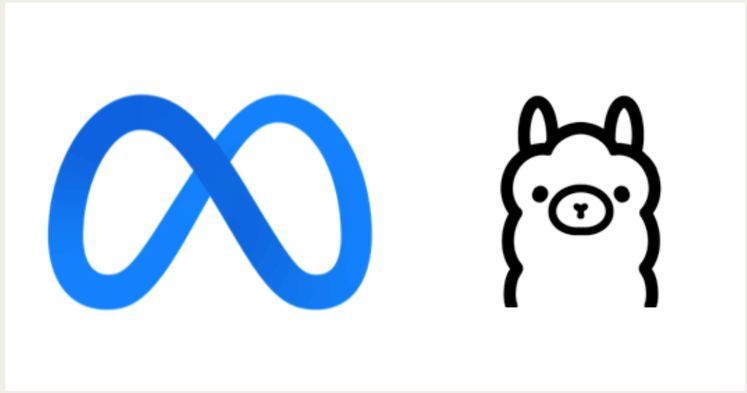
PROBLEM - WHAT DIDN'T WORK?



OLLAMA INSTALLED LOCAL LLMS DISPLAYED INCONSISTENT AND INCORRECT CATEGORIZATION EVEN WITH FEW SHOT PROMPTING.

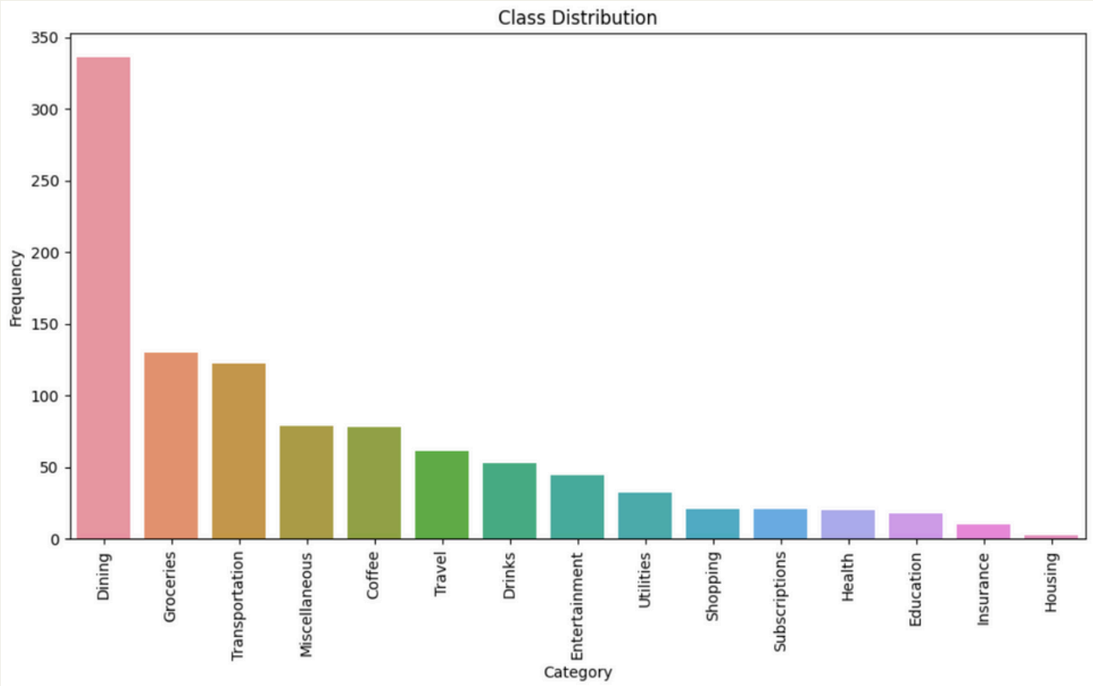
DATA

Date	Description	Amount
3/15/24	AplPay SPROUTS FARME DURHAM NC	15.38
3/15/24	BEAN TRADERS Durham NC	5.21
3/13/24	CHATGPT SUBSCRIPTION SAN FRANCISCO CA	21.8



- PROMPT: “YOU ARE A FINANCIAL ADVISOR. CAN YOU ASSIGN AN APPROPRIATE CATEGORY TO EACH TRANSACTION. MAINTAIN THE FORMAT: DATE - DESCRIPTION - AMOUNT - CATEGORY. CATEGORY NAMES SHOULD BE CONSISTENT AND LESS THAN 3 WORDS.”
- TEMPERATURE: 0.1
- MANUAL LABELING

DATA



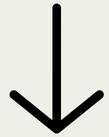
- 1226 ROWS IN TOTAL WITH TRANSACTION DATA FROM 2022-2024
- 15 CLASSES TO BEGIN WITH.
- IDENTIFIED CLASS IMBALANCE.
- LARGEST CLASS: DINING
- REMOVED CLASS WITH < 10 TRANSACTIONS - MISTAKES.
- REMOVED NEGATIVE TRANSACTIONS AS WANTED TO TRACK SPEND ONLY.

MODEL - NON DL APPROACH

TF-IDF + STOP WORDS

TRAIN-TEST SPLIT(20%)

RANDOM FOREST

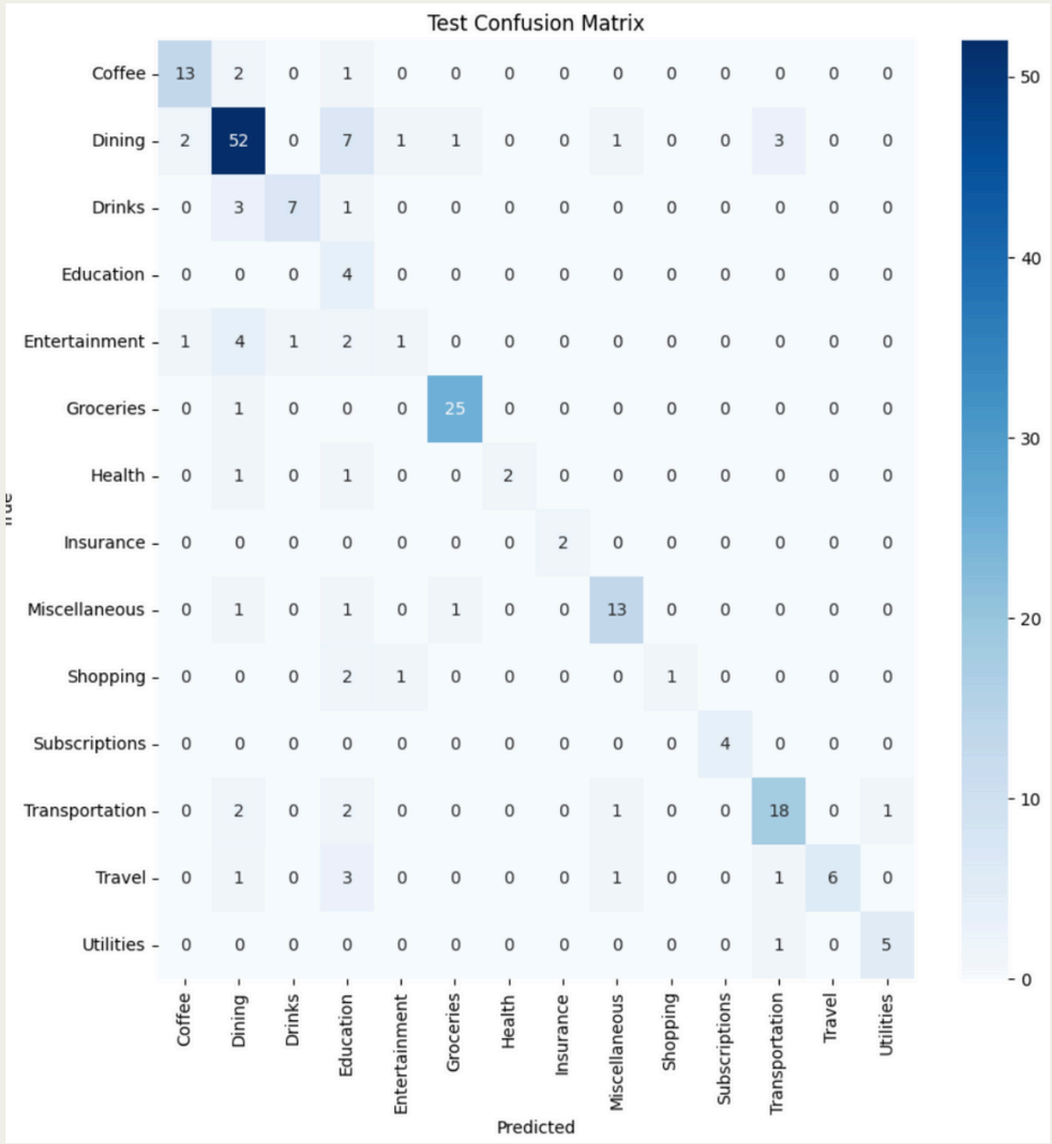


GRID SEARCH CV (5 FOLD)

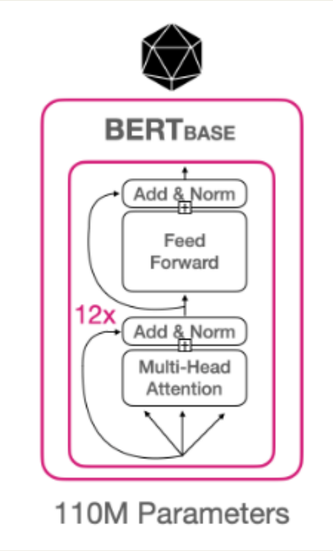
MODEL - RESULTS

IMPRESSIVE RESULTS OVERALL

- MODEL IS GETTING CONFUSED BETWEEN ENTERTAINMENT AND OTHER CATEGORIES.
- FOR NO NEURAL NETWORKS, NO GPU USE, AND FINE-TUNING PROCESS WHICH SHOWS THE POWER OF SIMPLE METHODS.
- CLASS IMBALANCE IS REAL BUT IS ALSO GOOD FOR THE MODEL TO LEARN AS A USER ATTRIBUTE.



MODEL - NAIVE DEEP LEARNING APPROACH



LABEL ENCODING

↓

TRAIN-TEST SPLIT (20%)

↓

BERT TOKENIZER - DESCRIPTION

↓

↓

PREDICTIONS

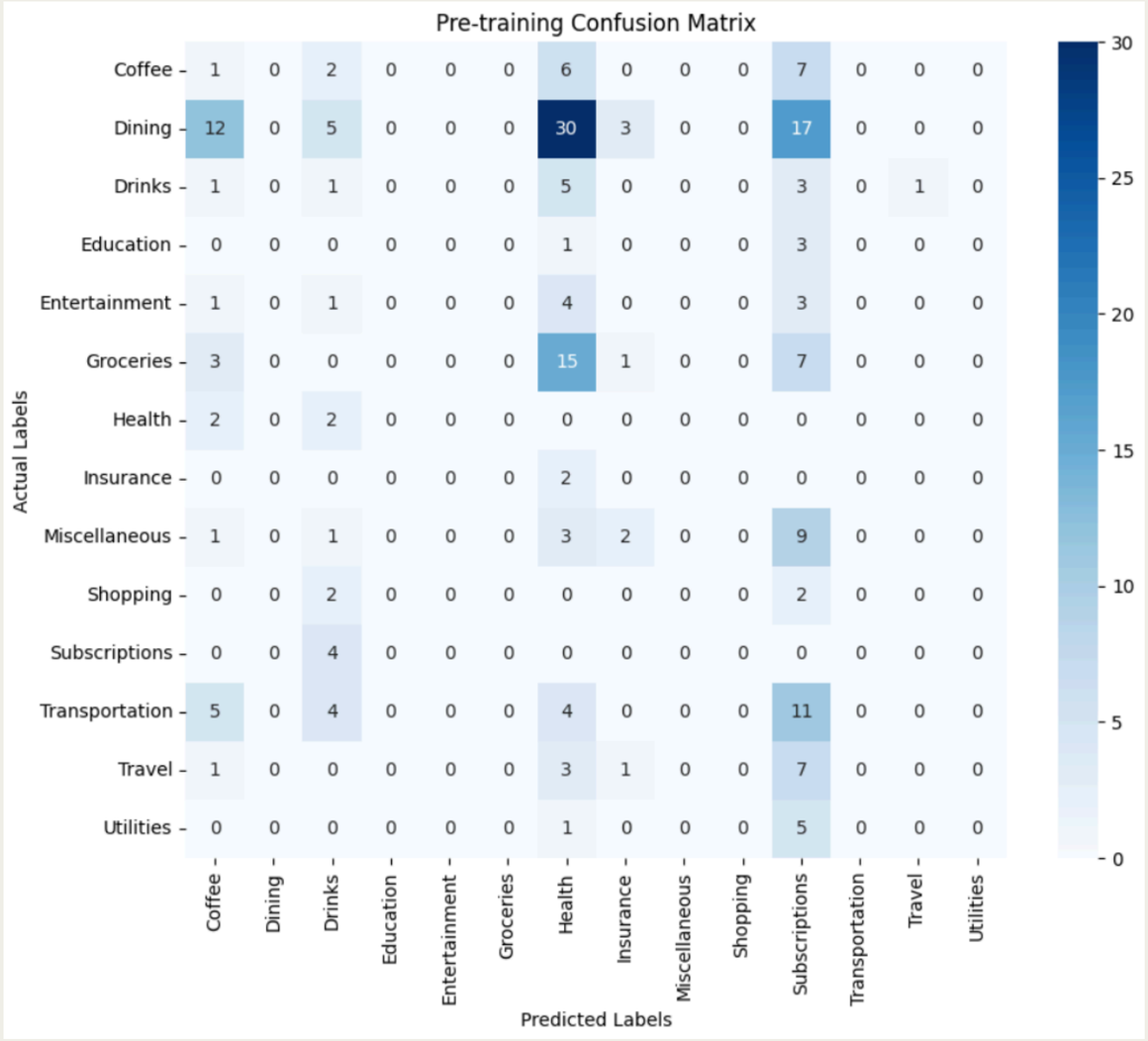
SIMPLE BERT-BASE UNCASED MODEL

- 110M PARAM
- UNCASED: CASE INSENSITIVE
- ALSO TRIED BERT-LARGE AND ROBERTA.

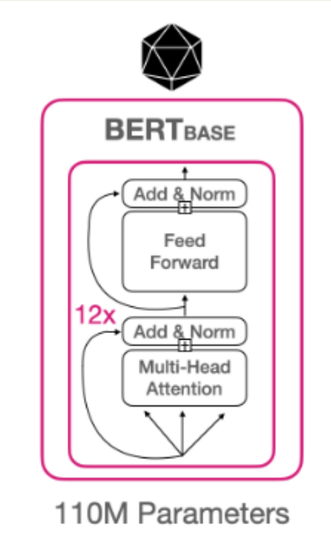
MODEL - RESULTS

POOR RESULTS OVERALL

- MODEL IS RANDOMLY PREDICTING CLASSES.
- MOST PREDICTIONS ARE SCATTERED BETWEEN HEALTH, SUBSCRIPTIONS, COFFEE, AND DRINKS.
- THIS BEHAVIOR WAS EXPECTED FOR BERT-BASE WITHOUT FINE-TUNING.



MODEL - DEEP LEARNING APPROACH



LABEL ENCODING

↓

TRAIN-TEST SPLIT

↓

BERT TOKENIZER - DESCRIPTION

↓

+

↓

PREDICTIONS

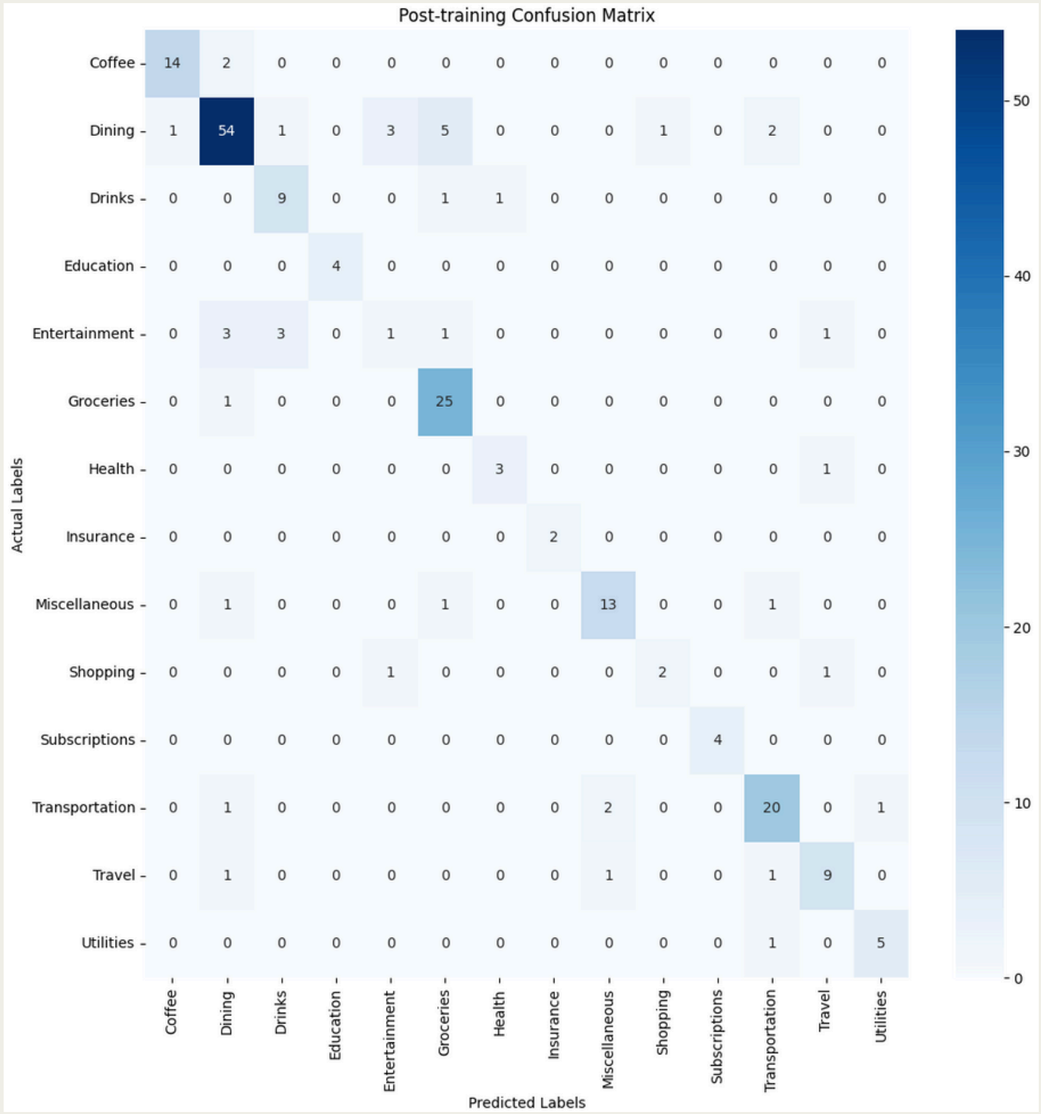
HUGGING FACE TRAINER

- FULL FINE-TUNING
- BATCH SIZE: 8
- LEARNING RATE: 5E-5
- OPTIMIZER: ADAMW
- WEIGHT DECAY: 0.01
- EPOCHS: 15

MODEL - RESULTS

BEST RESULTS OVERALL

- MODEL IS PREDICTING CLASSES ACCURATELY, WITH FEWER MISTAKES.
- FINE-TUNING TOOK ON 5 MINS.
- ENTERTAINMENT REMAINS THE CLASS OF BOTHER.
- A LARGER DATASET AND HYPERPARAMETER TUNING ARE SURE TO IMPROVE RESULTS.

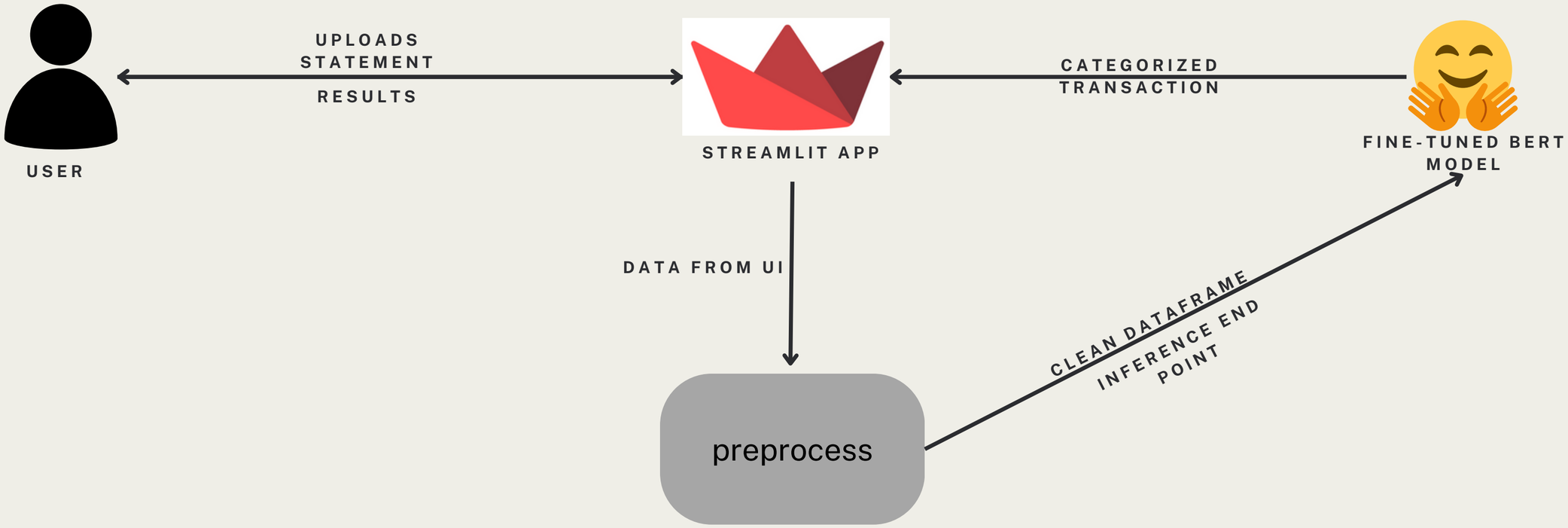


EVALUATION

Model	Accuracy	Precision	Recall	F-1 Score
TF-IDF + Random Forest	0.7463	0.8056	0.7463	0.7557
BERT - Naive	0.0098	0.0053	0.0098	0.0069
Fine-Tuned BERT	0.8049	0.7968	0.8049	0.7979

F-1 SCORE IS THE METRIC OF CHOICE DUE TO CLASS IMBALANCE AND EQUAL IMPORTANCE OF PRECISION AND RECALL.

PIPELINE


USER
UPLOADS STATEMENT
RESULTS
STREAMLIT APP
CATEGORIZED TRANSACTION
FINE-TUNED BERT MODEL
DATA FROM UI
preprocess
CLEAN DATAFRAME
INFERENCE END POINT

DEMO

CONCLUSIONS

- OUR PROOF OF CONCEPT ACHIEVED THE OBJECTIVE AT HAND.
- WE NOW HAVE A QUICK AND ULTRA-SECURE WAY TO ANALYZE PERSONAL FINANCES WITHOUT SHARING SENSITIVE DATA AT VERY LOW COST.
- PERFORMS WELL WITH A HIGH F-1 SCORE AND WITH MORE DATA CAN ACHIEVE PERFORMANCE EQUIVALENT TO MERCHANT CATEGORY CODINGS.
- STATE-OF-THE-ART LLMS LACK THE ABILITY TO CLASSIFY TRANSACTIONS CORRECTLY.
- ON A SMALLER DATASET, SIMPLE NON-DEEP LEARNING MODEL PERFORMED ALMOST AS GOOD AS THE FINE-TUNED BERT MODEL.

FUTURE GROWTH

- CURRENTLY ONLY FINE-TUNED TO MY TRANSACTIONS, ADD FUNCTIONALITY TO ALLOW USERS TO FINE-TUNE MODEL USING THEIR OWN BANK DATA.
- INTEGRATE ABILITY TO UPLOAD STATEMENTS FROM DIFFERENT BANKS.
- INCLUDE NEGATIVE TRANSACTIONS(CREDIT) TO SHOW INCOME.
- FIT A RAG BASED INSTRUCTION TUNED LLM TO GIVE PERSONALIZED FINANCIAL ADVICE BASED ON DATA.
- FINE-TUNE ON A LARGER DATASET TO IMPROVE F-1 SCORE.

Thank you!