AIML TASK OVERVIEW

Task 1

Image Tagging with TensorFlow

- **Objective:** Develop a model for image tagging using TensorFlow.
- **Data Preparation:** Gather a labeled dataset of images.
- **Model Architecture:** Design a CNN model for image classification.
- **Model Training:** Train the model using TensorFlow's Keras API.
- **Data Augmentation:** Enhance dataset with techniques like rotation, flipping.
- **Hyperparameter** Tuning: Optimize model performance with fine-tuning.
- **Model Evaluation:** Assess model accuracy, precision, recall.
- **Deployment:** Deploy the model for real-world image tagging applications.

Boston House Price Prediction

- UTILIZE THE PROVIDED DATASET FEATURING VARIOUS ATTRIBUTES INCLUDING THE NUMBER OF ROOMS, CRIME RATES, AND OTHER PERTINENT FACTORS.
- DESIGN AND IMPLEMENT A REGRESSION MODEL TO PREDICT BOSTON HOUSE PRICES ACCURATELY.
- THE SOLUTION INVOLVES SEVERAL KEY STEPS:
  A. DATA PREPROCESSING: CLEAN AND PREPROCESS THE DATASET, HANDLING MISSING VALUES AND OUTLIERS IF ANY.
  B. MODEL SELECTION: CHOOSE AN APPROPRIATE REGRESSION MODEL FOR THE PREDICTION TASK. OPTIONS MAY INCLUDE LINEAR REGRESSION, DECISION TREES, OR GRADIENT BOOSTING REGRESSORS.
  C. TRAINING: TRAIN THE SELECTED MODEL USING THE PREPROCESSED DATASET.
  D. EVALUATION: EVALUATE THE MODEL'S PERFORMANCE USING SUITABLE METRICS SUCH AS MEAN SQUARED ERROR (MSE) OR R-SQUARED (R2) SCORE TO ASSESS ITS PREDICTIVE ACCURACY.
  E. FINE-TUNING: FINE-TUNE THE MODEL PARAMETERS IF NECESSARY TO IMPROVE PERFORMANCE.
- THE GOAL IS TO DEVELOP A ROBUST REGRESSION MODEL THAT ACCURATELY PREDICTS BOSTON HOUSE PRICES, FACILITATING INFORMED DECISION-MAKING IN THE REAL ESTATE DOMAIN.

Autocorrect Keyboard System

OBJECTIVE: DEVELOP AN AUTOCORRECT KEYBOARD SYSTEM THAT IDENTIFIES MISSPELLED WORDS IN A TEXT INPUT AND PROVIDES AUTOCORRECTION SUGGESTIONS.

IMPLEMENTATION APPROACH:

UTILIZE THE SPELLCHECKER LIBRARY IN PYTHON TO IDENTIFY MISSPELLED WORDS AND PROVIDE AUTOCORRECTION SUGGESTIONS.

BENEFITS:

IMPROVES TEXT INPUT ACCURACY BY AUTOCORRECTING MISSPELLED WORDS IN REAL-TIME.
ENHANCES USER EXPERIENCE BY PROVIDING AUTOCORRECTION SUGGESTIONS, REDUCING MANUAL CORRECTION EFFORTS.

KEY COMPONENTS:

SPELLCHECKER: USED TO IDENTIFY MISSPELLED WORDS AND PROVIDE AUTOCORRECTION SUGGESTIONS.
EXPECTED OUTCOME: AN AUTOCORRECT KEYBOARD SYSTEM THAT IMPROVES TEXT INPUT ACCURACY BY AUTOMATICALLY CORRECTING MISSPELLED WORDS IN REAL-TIME, THEREBY ENHANCING USER EXPERIENCE DURING TYPING.

Task 2

Store Sales and Profit Analysis Task

THE STORE SALES AND PROFIT ANALYSIS TASK AIMS TO PROVIDE LEARNERS WITH THE NECESSARY KNOWLEDGE AND TOOLS TO ANALYZE RETAIL PERFORMANCE EFFECTIVELY USING PYTHON. BY FOCUSING ON SALES AND PROFITS, THIS TASK UNDERSCORES THE SIGNIFICANCE OF UTILIZING DATA TO DRIVE STRATEGIC BUSINESS DECISIONS. IT INVOLVES EXAMINING VARIOUS FACTORS SUCH AS SALES TRENDS, PROFITABILITY, PRODUCT CATEGORY PERFORMANCE, AND CUSTOMER SEGMENT CONTRIBUTIONS TO IDENTIFY OPPORTUNITIES FOR OPERATIONAL OPTIMIZATION, EFFECTIVE MARKETING STRATEGIES, AND INVENTORY MANAGEMENT IMPROVEMENTS.

KEY COMPONENTS OF THE TASK:
DATA PREPARATION:

CONVERTING DATE COLUMNS TO A DATETIME FORMAT TO EXTRACT VALUABLE TIME-BASED INSIGHTS.
HANDLING MISSING VALUES AND OUTLIERS TO PREPARE THE DATASET FOR ANALYSIS.
SALES ANALYSIS:

TEMPORAL SALES TRENDS: UNDERSTANDING MONTHLY SALES PATTERNS TO INFORM INVENTORY PLANNING AND PROMOTIONAL STRATEGIES.
CATEGORY AND SUB-CATEGORY ANALYSIS: IDENTIFYING TOP-PERFORMING PRODUCT CATEGORIES AND SUB-CATEGORIES TO PRIORITIZE STOCK AND REFINE PRODUCT OFFERINGS.

PROFIT ANALYSIS:

PROFIT TRENDS: IDENTIFYING THE MOST PROFITABLE PRODUCTS TO OPTIMIZE PRODUCT MIX.
SEGMENTATION ANALYSIS: EVALUATING PROFITABILITY BY CUSTOMER SEGMENT TO TAILOR MARKETING STRATEGIES EFFECTIVELY.
VISUALIZATION:

UTILIZING PYTHON'S PLOTLY LIBRARY FOR DYNAMIC AND INTERACTIVE VISUALIZATIONS TO MAKE DATA INSIGHTS ACCESSIBLE AND ACTIONABLE.

OPERATIONAL INSIGHTS:

SALES-TO-PROFIT RATIOS: ASSESSING EFFICIENCY AND PROFITABILITY ACROSS DIFFERENT SEGMENTS OR PRODUCT CATEGORIES.
CUSTOMER SEGMENT PERFORMANCE: EVALUATING THE PROFITABILITY OF CUSTOMER SEGMENTS TO FOCUS EFFORTS AND RESOURCES EFFECTIVELY.

Car Price Prediction

PROBLEM DEFINITION:

CLEARLY DEFINE THE PROBLEM STATEMENT, WHICH IS TO PREDICT CAR SELLING PRICES BASED ON VARIOUS ATTRIBUTES.

DATA COLLECTION:

GATHER A DATASET CONTAINING INFORMATION ABOUT CARS, INCLUDING ATTRIBUTES LIKE FUEL TYPE, YEARS OF SERVICE, SHOWROOM PRICE, PREVIOUS OWNERS, KILOMETERS DRIVEN, SELLER TYPE, AND TRANSMISSION TYPE.

DATA PREPROCESSING:

HANDLE MISSING VALUES, REMOVE IRRELEVANT COLUMNS, DERIVE NEW FEATURES (E.G., YEARS SINCE MANUFACTURE), AND ENCODE CATEGORICAL VARIABLES.

EXPLORATORY DATA ANALYSIS (EDA):

ANALYZE THE DATASET TO UNDERSTAND THE DISTRIBUTIONS, CORRELATIONS, AND RELATIONSHIPS BETWEEN VARIABLES. THIS STEP HELPS IN IDENTIFYING PATTERNS AND INSIGHTS.

FEATURE ENGINEERING:

EXTRACT RELEVANT FEATURES, TRANSFORM VARIABLES IF NEEDED, AND PREPARE THE DATA FOR MODEL TRAINING.

DATA SPLITTING:

SPLIT THE DATASET INTO TRAINING AND TEST SETS TO EVALUATE THE PERFORMANCE OF THE MODEL.
MODEL SELECTION:

CHOOSE AN APPROPRIATE MACHINE LEARNING ALGORITHM FOR REGRESSION TASKS. IN THIS CASE, A RANDOM FOREST REGRESSOR IS SELECTED DUE TO ITS ABILITY TO HANDLE COMPLEX DATA AND PROVIDE ACCURATE PREDICTIONS.
HYPERPARAMETER TUNING:

OPTIMIZE THE HYPERPARAMETERS OF THE SELECTED MODEL USING TECHNIQUES LIKE RANDOMIZEDSEARCHCV TO IMPROVE ITS PERFORMANCE.

MODEL TRAINING:

TRAIN THE SELECTED MODEL USING THE TRAINING DATASET TO LEARN PATTERNS AND RELATIONSHIPS BETWEEN INPUT FEATURES AND TARGET VARIABLE (SELLING PRICES).
MODEL EVALUATION:

EVALUATE THE TRAINED MODEL'S PERFORMANCE USING THE TEST DATASET, MEASURING METRICS LIKE MEAN SQUARED ERROR (MSE) OR ROOT MEAN SQUARED ERROR (RMSE) TO ASSESS PREDICTION ACCURACY.
DEPLOYMENT:

DEPLOY THE TRAINED MODEL AS A WEB APPLICATION OR API USING FRAMEWORKS LIKE FLASK, ENABLING USERS TO INPUT CAR DETAILS AND RECEIVE ESTIMATED SELLING PRICES.
CONTINUOUS MONITORING AND IMPROVEMENT:

MONITOR THE DEPLOYED MODEL'S PERFORMANCE AND GATHER FEEDBACK TO IDENTIFY AREAS FOR IMPROVEMENT. ITERATIVELY REFINE THE MODEL BASED ON NEW DATA AND USER REQUIREMENTS TO ENSURE ITS EFFECTIVENESS OVER TIME.

LOAN APPROVAL PREDICTION WITH MACHINE LEARNING

Task Overview: Loan Approval Prediction with Machine Learning

Objective:
This task aims to develop a machine learning model capable of predicting loan approvals based on applicants' credit history data. By employing Python and libraries like TensorFlow or PyTorch, interns will have to create an intelligent system to classify loan applications into approved or denied.

Dataset:
An ideal dataset for this project has been sourced, containing historical loan application data. Interns will analyze various factors including financial history, income, credit rating, and employment status.

Methodology:

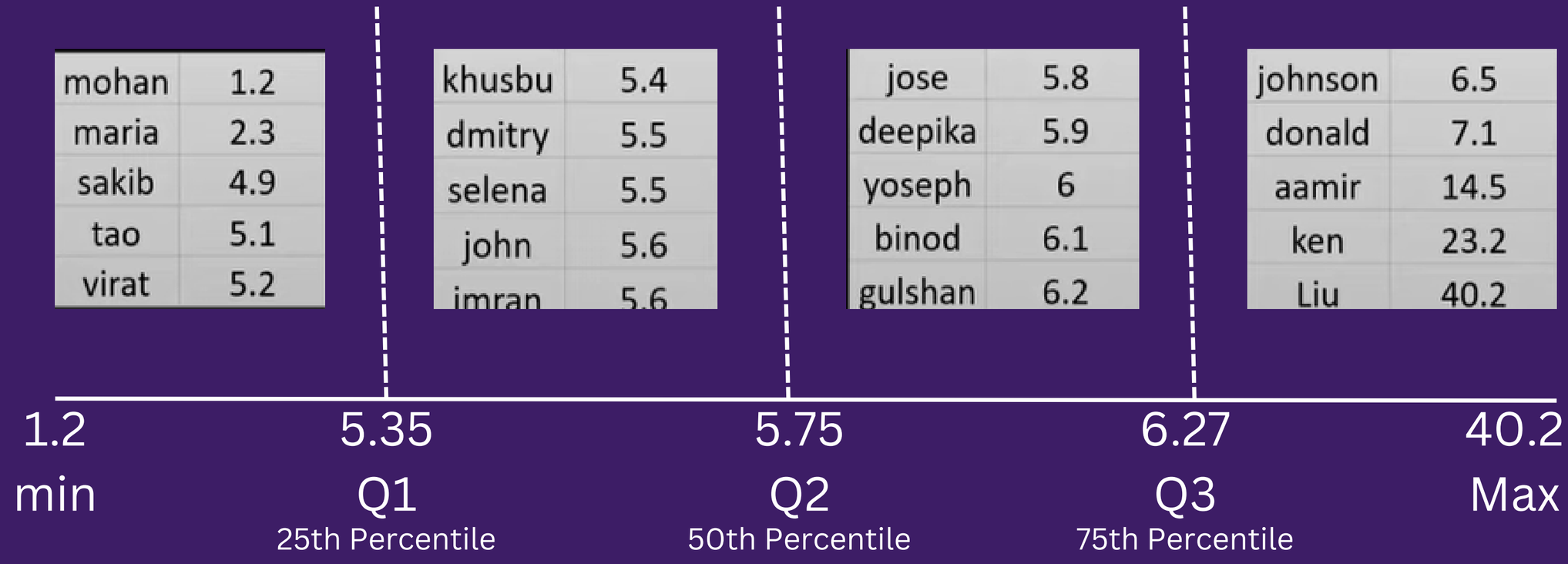
Data Preprocessing: Begin by importing necessary libraries (pandas, numpy, sklearn) and loading the dataset. Clean the data by dropping irrelevant columns, filling missing values with appropriate statistics (mode for categorical and median/mode for numerical values), and handling outliers.
Exploratory Data Analysis (EDA): Utilize libraries like plotly for visualizing the distribution of loan statuses, gender, marital status, education, and other variables to uncover insights.
Feature Engineering: Convert categorical variables into numerical using one-hot encoding, preparing the data for model training.
Model Training: Train a loan approval prediction model using the RandomForestClassifier or SVC from sklearn. Experiment with different algorithms if necessary to improve accuracy.
Evaluation: Split the data into training and testing sets, scale the features, and evaluate the model's performance on the test set.

Expected Outcomes:

A functional machine learning model that accurately predicts loan approval status.
A comprehensive report documenting the methodology, code snippets, and analysis of model performance.

IQR Technique Example

name	height
mohan	1.2
maria	2.3
sakib	4.9
tao	5.1
virat	5.2
khusbu	5.4
dmitry	5.5
selena	5.5
john	5.6
imran	5.6
jose	5.8
deepika	5.9
yoseph	6
binod	6.1
gulshan	6.2
johnson	6.5
donald	7.1
aamir	14.5
ken	23.2
Liu	40.2



IQR=Q3-Q1
=6.7-5.35
=0.925

Lower limit=Q1-1.5*IQR
=3.96
Upper limit=Q361.5*IQR
=7.66

Outliers are Excluded

name	height
mohan	1.2
maria	2.3
sakib	4.9
tao	5.1
virat	5.2
khusbu	5.4
dmitry	5.5
selena	5.5
john	5.6
imran	5.6
jose	5.8
deepika	5.9
yoseph	6
binod	6.1
gulshan	6.2
johnson	6.5
donald	7.1
aamir	14.5
ken	23.2
Liu	40.2

Task 3

AI-DRIVEN NATURAL LANGUAGE PROCESSING PROJECT

LM Selection: Choose an LM aligned with interests, whether GPT-3, BERT, or specialized models.
Implementation: Create a Jupyter notebook to showcase the chosen LM's implementation, emphasizing unique features.
Exploration and Analysis: Thoroughly explore the LM's capabilities, analyzing performance on various inputs.
Research Questions: Define research questions exploring strengths and limitations, tailored to extract meaningful insights.
Project Alignment: Align project with NLP and ML advancement goals, considering ethical considerations.
Conclusion and Insights: Summarize findings, draw insightful conclusions, and discuss potential applications or areas for improvement.

Langchain

Your LLm Application

LangChain

Open AI

Hugging Face

...

THANK YOU