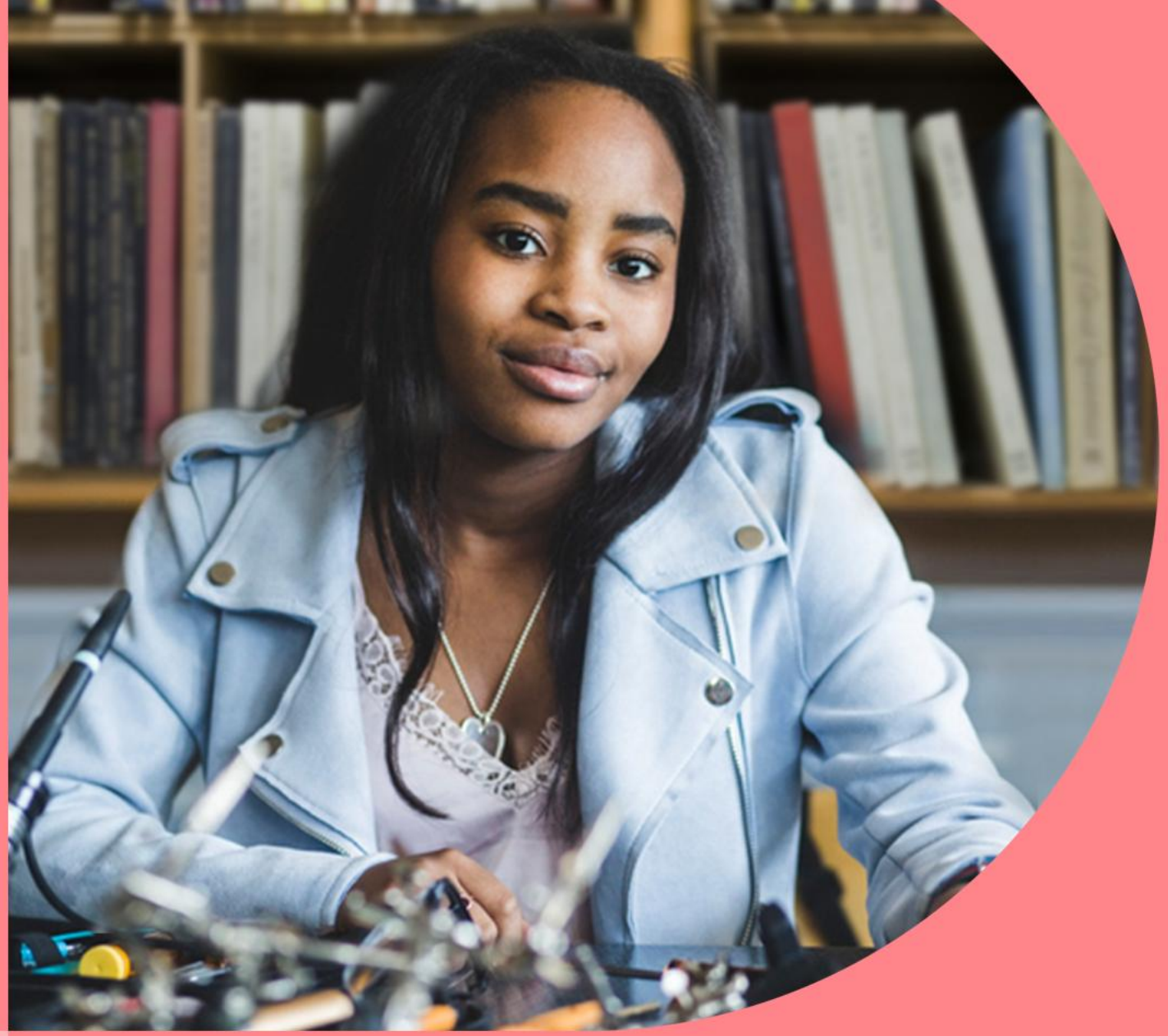


AI for Youth

Project 400 Local Deployment of AI Project & Presentation

intel digital readiness

Legal Disclaimers

The Intel® Digital Readiness Programs and Intel® AI for Youth program are developed by Intel Corporation.

Intel technologies may require enabled hardware, software, or service activation.

No product or component can be absolutely secure.

Results have been estimated or simulated.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

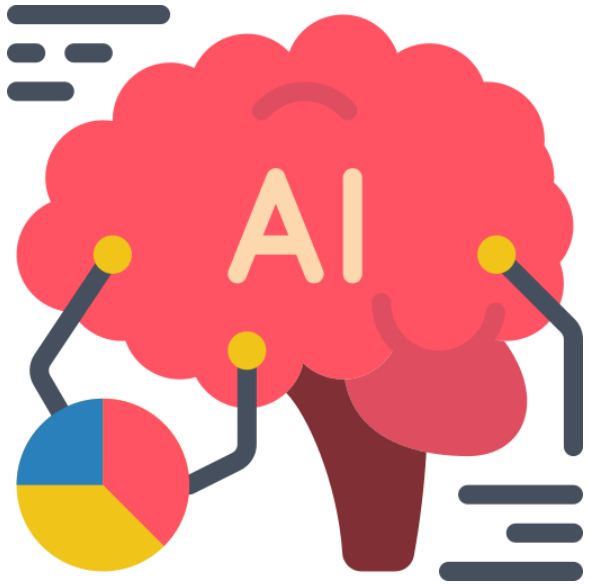
Activity: Project Presentation

Duration: 360 Minutes

intel digital readiness

Activity Introduction

- In this activity, youth will be able to create an AI Project on their own based on the concepts learned so far.
- Youth will also be able to build the project using AI Project Cycle, following the given guidelines and template.
- Additionally, youth will be deploying their project using Streamlit.



Activity Guidelines

- Follow the template and fill in the relevant sections at the relevant time intervals
- The project presentation, which is the last part, will be used for the presentation
- Some parts of the presentation part will be repetitive, in which case youth can directly copy the templates they have already filled

Activity: Building the AI problem statement

50 minutes

intel digital readiness

Think about problems you observe everyday

Write down some problems you have noticed in everyday life
(Hint: It could be related to the environment, society, community, school, family, etc.)

1. Road traffic accidents due to reckless driving and speeding.
2. Poor waste management and littering in urban areas.
3. Illegal mining (“galamsey”) causing environmental destruction.
4. Flooding in Accra due to poor drainage.

List down the problems in the descending order of social impact
(Problems with the highest social impact come first)

1. Road traffic accidents (human lives at stake).
2. Illegal mining destroying water bodies.
3. Flooding displacing communities.
4. Poor waste management (public health impact).

Another way of coming up with a problem statement

List down some problems you think can be addressed from the SDG targets list
(Go to this link, https://www.globalgoals.org/goals, where you can find out about the targets of the selected SDGs and choose one or more targets and come up with related problems)

1. Youth Unemployment → SDG 8 (Decent Work and Economic Growth) and SDG 10 (Reduced Inequalities).

2. Illegal mining → SDG 6 (Clean Water & Sanitation) and SDG 15 (Life on Land).

3. Flooding → SDG 13 (Climate Action) and SDG 11 (Sustainable Cities & Communities).

4. Waste management → SDG 12 (Responsible Consumption & Production).

List down all the problems which you witnessed every day which can be connected to the SDGs

1. Road traffic accidents
2. Illegal mining
3. Youth Unemployment
4. Waste Management

Which of these problems can be solved with AI?

1. Youth Unemployment
2. Illegal mining
3. Flooding
4. Waste Management

List down all the resources that you find

The solution	Links	Authors
Highlights accident hotspots and causes	https://ama.gov.gh/documents/2023_Accra_Road_Safety_Report_final.pdf	National Road Safety Authority (NRSA)
Statistical analysis of accident patterns	https://doi.org/10.30654/MJEM.10048	Nathaniel Gyimah
Detects risky driver actions from camera feed	https://docs.ultralytics.com/modes/track/	Ultralytics
Pre-trained models for object detection and behavior recognition	https://docs.openvino.ai/2024/documentation/legacy-features/model-zoo.html	OpenVINO Model Zoo
AI-powered road safety monitoring systems	https://smartcitymall.africa/ai-traffic-management-in-smart-cities/	SmartCity Africa

4Ws Problem Scoping Template

Difficulty in Matching Ghanaian Youth to Relevant Job Opportunities

Our	primary stakeholders are Ghanaian youth especially recent graduates	Who
have a problem that	struggle to find jobs because postings are scattered, uncategorized, and often unclear	What
when / while	searching through multiple online job portals, social media groups, and notice boards	Where
An ideal solution would	classify job descriptions into clear categories and make job search faster and more targeted—leading to higher employment rates and reduced application frustration	Why

Create a problem statement

Choose one problem with the highest social impact and describe it briefly

Many Ghanaian youth struggle to find jobs that match their skills because job postings are scattered, poorly categorized, and inconsistent in format. This forces job seekers to spend excessive time filtering through irrelevant opportunities, leading to frustration and missed chances.

An AI-powered job classification and delivered through a simple web interface, can streamline job searches, improve relevance, and increase employment opportunities for young people.

Conversion into an AI problem statement

Think about how we can solve the problem using AI and write an AI problem statement

1. How can AI be used to solve the problem? (mention the solution to the problem)

AI can automatically analyse messy job descriptions and classify them into structured categories.

2. Which domains of AI is the problem related to?

- ☐ Computer Vision
- ☐ Statistical Data
- ☒ Natural Language Processing

3. Write down the AI problem statement in one line

To develop an AI-powered system that classifies unstructured job postings and delivers personalized recommendations to Ghanaian youth for speed and inclusivity.

Activity: Search and find the dataset

50 minutes

intel digital readiness

Try to list down the sources of datasets for the resources listed earlier

Title	Author	Date	Website Link
Applied Text Analysis with Python	Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda	2018	https://www.oreilly.com/library/view/applied-text-analysis/9781491963036/
Classifying Jobs as Fraudulent or Real Using TF-IDF, Logistic Regression, and a Random Forest Classifier	Kai Michael Dalen	2023	https://medium.com/@kai.michael.dalen/classifying-jobs-as-fraudulent-or-real-using-tf-idf-logistic-regression-and-a-random-forest-22af3535b541
Unlocking Potential: A Machine Learning Approach to Job Category Prediction	Swati Garg, Chandra Sekhar, Lov Kumar	2024	https://www.researchgate.net/publication/385955873_Unlocking_Potential_A_Machine_Learning_Approach_to_Job_Category_Prediction
OpenVINO Toolkit	Intel	2023	https://www.intel.com/content/www/us/en/developer/tools/openvino-toolkit/overview.html

Activity: Perform Data Exploration on the dataset

50 minutes

intel digital readiness

Checklist - Computer Vision

Data exploration and Pre-processing

- ❑ Checking the blurriness of the images
- ❑ Checking the colour distribution
- ❑ Checking for the class imbalance

Pre-processing

- ❑ Checking the size of the images
- ❑ Resizing
- ❑ Reshaping the image
- ❑ Detecting the edges

Try to use this as a checklist based on your use-case.

Note: You may not require all the steps mentioned in this checklist for your project.

Checklist - Statistical data

Data Exploration

- ❑ Check for outliers in the data
- ❑ Check the distribution of data
- ❑ Check Data Skewness
- ❑ Check for missing values in the data
- ❑ Check for correlation
- ❑ Check for patterns in the dataset by visualization

Pre-processing

- ❑ Outliers Treatment
- ❑ Missing value Treatment
- ❑ Dealing with correlation
- ❑ Feature engineering
- ❑ Scaling
- ❑ Separating dependent and independent variables
- ❑ Train-test split

Try to use this as a checklist based on your use-case.

Note: You may not require all the steps mentioned in this checklist for your project.

Checklist - Natural Language Processing

Data exploration

- ❑ Visualize the frequency of words
- ❑ Check Corpus length
- ❑ Check Document length
- ❑ Check Sentence length
- ❑ Average word length analysis
- ❑ Stop words analysis

Pre-processing

- ❑ Cleaning the data
- ❑ Tokenization
- ❑ Removing stop words
- ❑ Lemmatization and Stemming
- ❑ Creating document term matrix
- ❑ Converting to the bag of words

Try to use this as a checklist based on your use-case.

Note: You may not require all the steps mentioned in this checklist for your project.

Activity: Choose and create the model that solves the problem

50 minutes

intel digital readiness

Choose the right model

- Based on the type of data and the problem statement you have acquired, choose the correct model
- Please refer to the checklists for each domain in the following slides
- What type of model have you chosen?

☒ Supervised Learning ☐ Unsupervised Learning

Computer Vision

Note: You may not require all the steps mentioned in this checklist for your project.

ML Models

- ❑ Logistic Regression
- ❑ Random Forest
- ❑ CNN
- ❑ Other ML models

Popular pre-trained models

- ❑ VGG-16 – OpenVINO Toolkit
- ❑ AlexNet – OpenVINO Toolkit
- ❑ ResNet-50 – OpenVINO Toolkit
- ❑ EfficientNet – OpenVINO Toolkit

Statistical Data

Note: You may not require all the models mentioned in this checklist for your project.

Popular models: Regression

- ❑ Linear Regression
- ❑ Lasso Regression
- ❑ Decision Tree Regression
- ❑ Random Forest
- ❑ KNN Model
- ❑ Support Vector Machines (SVM)

Popular models: Classification

- ❑ Logistic Regression
- ❑ Naive Bayes
- ❑ K-Nearest Neighbors
- ❑ Decision Tree
- ❑ Support Vector Machines

Natural Language Processing

Note: You may not require all the steps mentioned in this checklist for your project.

ML Models

- ❑ Logistic Regression
- ❑ Random Forest
- ❑ Other ML model

Popular pre-trained models

- ❑ BERT
- ❑ GPT 2
- ❑ Roberta_base
- ❑ DistilBERT

Refer to the Hugging Face library to use these models:
https://huggingface.co/models

Activity: Perform Evaluation of the chosen model

50 minutes

intel digital readiness

Which evaluation metric is the appropriate for your use case?

- Which metrics do you need to use?

1. Accuracy
2. The F1-Macro
3. The F1-Weighted metric
4. The Classification Report
5. A Confusion Matrix

Which evaluation metric is the appropriate for your use case?

Evaluation metrics: Regression

- ❑ MSE
- ❑ RMSE
- ❑ Mean Absolute Error
- ❑ R Squared (R2)

To learn more about these metrics, go to this link:

- Link 1: https://www.freecodecamp.org/news/evaluation-metrics-for-regression-problems-machine-learning

Evaluation metrics: Classification

- ❑ Confusion Matrix
- ❑ F1 Score
- ❑ Log Loss
- ❑ Categorical Cross entropy
- ❑ AUC

To learn more about these metrics, go to these link:

- Link 1: https://www.kdnuggets.com/2020/04/performance-evaluation-metrics-classification.html
- Link 2: https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification

Activity: Deploy the model using Streamlit

50 minutes

intel digital readiness

Which Streamlit features did you use for your use application?

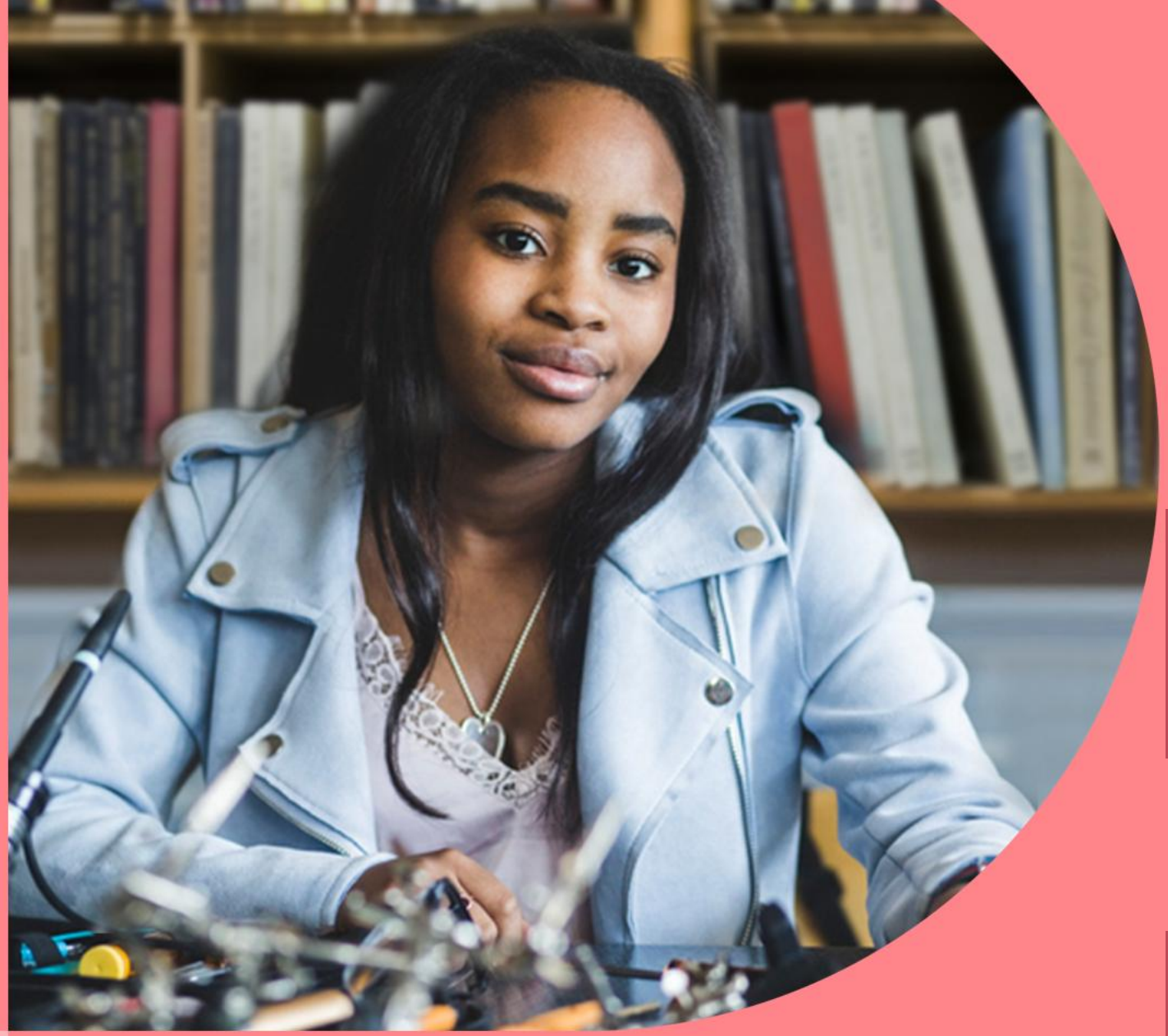
- Which are the features you used in Streamlit?

- st.title() / st.header() / st.subheader()
- st.text_input() / st.text_area()
- st.button()
- st.write() / st.markdown()
- st.success() / st.warning() / st.error() / st.info()
- st.spinner()

Activity: Project presentation

50 minutes

intel digital readiness



AI for Youth 2.0

Project name:

_____

_____

Names of team members:

_____

_____

intel digital readiness

AI Problem Statement:

Youth unemployment in Ghana is worsened by the disorganized and scattered nature of job postings, which are spread across various platforms and often have inconsistent descriptions. This makes it hard for young job seekers, especially recent graduates, to find relevant opportunities.

This project aims to solve this problem by developing an **AI-powered job classification system**. This system will organize job descriptions into structured categories through an accessible web interface. By making the job search more efficient, the project intends to empower Ghanaian youth, enhance job-matching outcomes.

Create a problem statement

Many Ghanaian youth struggle to find jobs that match their skills because job postings are scattered, poorly categorized, and inconsistent in format. This forces job seekers to spend excessive time filtering through irrelevant opportunities, leading to frustration and missed chances.

An AI-powered job classification and delivered through a simple web interface, can streamline job searches, improve relevance, and increase employment opportunities for young people.

4Ws Problem Scoping Template

Difficulty in Matching Ghanaian Youth to Relevant Job Opportunities

Our	primary stakeholders are Ghanaian youth especially recent graduates	Who
have a problem that	struggle to find jobs because postings are scattered, uncategorized, and often unclear	What
when / while	searching through multiple online job portals, social media groups, and notice boards	Where
An ideal solution would	classify job descriptions into clear categories and make job search faster and more targeted—leading to higher employment rates and reduced application frustration	Why

List the resources used to build this project

Title	Author	Date	Website Link
Applied Text Analysis with Python	Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda	2018	https://www.oreilly.com/library/view/applied-text-analysis/9781491963036/
Classifying Jobs as Fraudulent or Real Using TF-IDF, Logistic Regression, and a Random Forest Classifier	Kai Michael Dalen	2023	https://medium.com/@kai.michael.dalen/classifying-jobs-as-fraudulent-or-real-using-tf-idf-logistic-regression-and-a-random-forest-22af3535b541
Unlocking Potential: A Machine Learning Approach to Job Category Prediction	Swati Garg, Chandra Sekhar, Lov Kumar	2024	https://www.researchgate.net/publication/385955873_Unlocking_Potential_A_Machine_Learning_Approach_to_Job_Category_Prediction
OpenVINO Toolkit	Intel	2023	https://www.intel.com/content/www/us/en/developer/tools/openvino-toolkit/overview.html

Data Source

What is the source of the data used in the project?

The source of the data used in this project is the Jobberman Ghana website (https://www.jobberman.com.gh).

name	job_url	hiring_firn	hiring_firn	job_functi	title	date_post	Location	Job type	Industry	Salary	details	info	job_descr	requireme	responsib	qualificati	benefits	company_	applicatic	emplo
Executive	https://www.jobberman.com.gh				Executive Recruitment			Full Time			['Experien	['Job Title:	placeholder	Explore and	disco	manufacturing & warehousing remote (work fro				
Executive I	https://www.jobberman.com.gh/job-seek				Executive Recruitment			Full Time			['Experien	['Job Title:	placeholder	Discove	s easily.	job alerts and one-click applications sign up for j				
Executive I	https://www.jobberman.com.gh/employe				Executive Recruitment			Full Time			['Experien	['Job Title:	placehold	, easy one	. easily filt	and brain	solutions designed to take your recru			
Vehicle M	https://ww	Confidenti	https://www.jobbern		Vehicle Maintenance		Accra & Te	Part Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	conduct regular insp	â€“ on call allowance contract type: p				
Marketing	https://ww	Confidenti	https://www.jobbern		Marketing and Sales		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	identify and target ne	competitive base salary + uncapped c				
Security O	https://ww	Confidenti	https://www.jobbern		Security Officer		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	patrol pre	wassce or equivalent required.proven experienc				
Sales Exec	https://ww	Votila Bio-	https://www.jobbern		Sales Executive		Kumasi &	Full Time		Confidenti	['Experien	['Job Title:	New Featu	basic job description	/diploma in supply chain/marketing/agribusine					
Area Sales	https://ww	Jobberma	https://www.jobbern		Area Sales Executive		Area Sales	Full Time		Confidenti	['Experien	['Job Title:	New Featu	responsib	generate p	bachelors in marketing or business admi	full-ti			
Area Sales	https://ww	Jobberma	https://www.jobbern		Area Sales Executive		Area Sales	Full Time		Confidenti	['Experien	['Job Title:	New Featu	responsib	generate p	bachelors in marketing or business admi	full-ti			
Field Recr	https://ww	Cocoasis	https://www.jobbern		Field Recruitment Le		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	responsib	in this role	possess good grammatical skills and be a gooc				
Accounts	https://ww	Stallion Ti	https://www.jobbern		Accounts Receivable		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	accounts	experience level : entry level experience length :				
We are loc	https://www.jobberman.com.gh/jobs?q=:				We are looking for ACCOUNTS F			Full Time			['Experien	['Job Title:	placeholder	Find a Job + + + +	Search Search Filter Homepage Search results Jol					
Data Clerk	https://ww	Anonymou	https://www.jobbern		Data Clerk/Engineeri		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	job descri	s within th	experience level : mid level experience len	full ti			
Sales Exec	https://ww	Reputable	https://www.jobbern		Sales Executives (Ku		Sales Exec	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	pursue sa	experience level : entry level experience length :				
Senior Ma	https://ww	Confidenti	https://www.jobbern		Senior Marketing Ma		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	qualities e	.health ins	background: marketing or business degre	full-ti			
Sales Mar	https://ww	Anonymou	https://www.jobbern		Sales Manager		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	plan and c	experience level : senior level experience l	full-ti			
Engineer/	https://ww	Resolve Sy	https://www.jobbern		Engineer/ Technician		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	Immediat	job descri	include in	and experience: bachelorâ€™s degree or higher				
Regional M	https://ww	Confidenti	https://www.jobbern		Regional Manager		Regional M	Full Time		Confidenti	['Experien	['Job Title:	New Featu	main duti	oversees a	experience level : mid level experience len	full ti			
Sales Advi	https://ww	Petra Trus	https://www.jobbern		Sales Advisor		Accra & Te	Full Time		Confidenti	['Experien	['Job Title:	New Featu	key respor	drive sale	bachelorâ€™s degree in business administratio				
School Pri	https://ww	Kester Inte	https://www.jobbern		School Principal (Ac		School Pri	Full Time			['Experien	['Job Title:	New Featu	key respor	provide st	and experience: bachelorâ€™s degree in e	full-ti			

ghana_jobs_detailed

Paste the images/screenshots here

Describing the Data

Attributes of the data

```
(6, 26)
Index(['name', 'job_url', 'hiring_firm', 'hiring_firm_url', 'job_function',
       'title', 'date_posted', 'Location', 'Job type', 'Industry', 'Salary',
       'details', 'info', 'job_description', 'requirements',
       'responsibilities', 'qualifications', 'benefits', 'company_info',
       'application_deadline', 'employment_type', 'experience_required',
       'education_required', 'skills_required', 'salary_range',
       'location_details'],
      dtype='object')
                                    name  \
0                     Executive Recruitment
1                     Executive Recruitment
2                     Executive Recruitment
3  Vehicle Maintenance and Appearance Officer
4                Marketing and Sales Officer

                                        job_url   hiring_firm  \
0                     https://www.jobberman.com.gh          NaN
1           https://www.jobberman.com.gh/job-seeker          NaN
2             https://www.jobberman.com.gh/employer          NaN
3  https://www.jobberman.com.gh/listings/vehicle-...  Confidential
4  https://www.jobberman.com.gh/listings/marketin...  Confidential

                                hiring_firm_url  job_function  \
```

Paste the images/screenshots here

Data Exploration

- **Step 1:** I began by checking the dataset's shape (rows, columns) and identifying the count of non-null values for each column. This made me have a quick assessment of data completeness. I explored the distribution of key categorical and numerical features where I used a bar chart to show the distribution of jobs across different categories, revealing the most common job types.

- **Step 2:** I did a word frequency analysis. I identified the most frequent words overall and presented them in a bar chart. I refined the word frequency analysis by grouping by category, identifying the most common and often discriminative words for each specific job type.

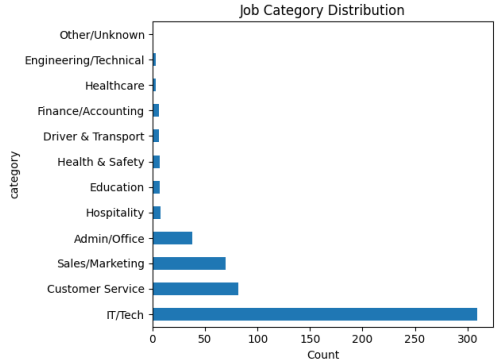
- **Step 3:** I went beyond single words to find common phrases. I specifically analyzed: Bigrams (two-word phrases) like "project manager" or "sales executive". Trigrams (three-word phrases) like "minimum years experience" or "strong communication skills". This helps to reveal multi-word concepts and requirements.

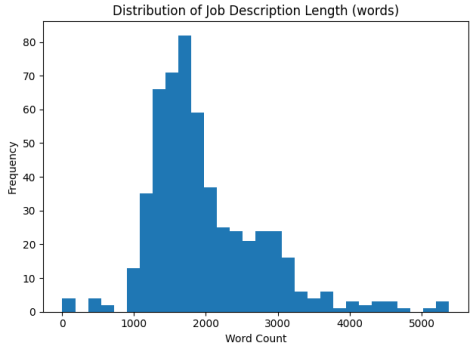
Data Exploration – Step 1

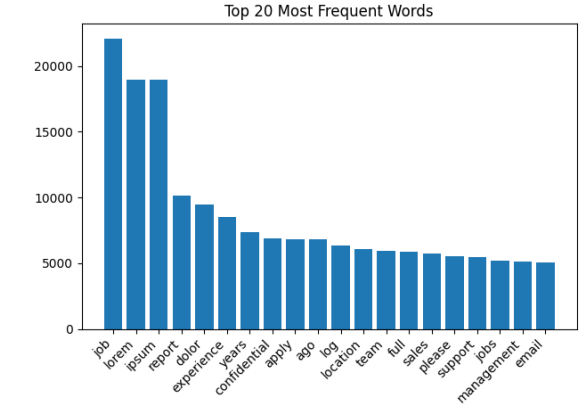
Step 1:

```
==================
1) Dataset Summary
==================
Shape: (540, 32)
Columns: ['name', 'job_url', 'hiring_firm'

Non-null counts:
name                    540
job_url                 540
title                   540
Job type                540
details                 540
info                    540
category                540
desc_len                540
clean_text              540
raw_text                540
job_description         536
skills_required         536
requirements
Location                512
hiring_firm             495
hiring_firm_url         495
Salary                  494
```







Paste the images/screenshots here

Observations from Step 1

It tells me this dataset has over 500+ rows. From the images, we could see that the most prominent category is IT/Tech which accounts for over half of all jobs with 309 listings. The job description lengths vary widely with an average of about 1,985 words. The histogram shows a bell-shaped curve, indicating that most descriptions are clustered between 1,500 and 2,500 words.

Modeling

Enlist the ML or DL models used below:

1. Model 1: TF-IDF + Cosine Similarity
2. Model 2: TF-IDF + Logistic Regression
3. Model 3: DistilBERT (MiniLM) Embeddings

Describe the best ML models used:

1. Model 1: TF-IDF + Cosine Similarity: It uses TfidfVectorizer to convert job descriptions into numerical vectors, where each value represents the importance of a word in a document relative to the entire corpus. To classify a new job, the model finds the most similar job in the training set using cosine similarity (the cosine of the angle between two vectors). It then simply assigns the category of that closest job to the new job.
2. Model 2: TF-IDF + Logistic Regression: It then trains a Logistic Regression classifier on these vectors. This model learns the relationship between the presence of certain words (represented by TF-IDF scores) and the target job category. It also uses class weights (class_weight='balanced') to address the imbalanced class distribution, ensuring the model doesn't ignore the less common job categories.
3. Model 3: DistilBERT (MiniLM) Embeddings: It uses a pre-trained Sentence Transformer model (all-MiniLM-L6-v2) to convert each job description into a dense numerical vector (an embedding). These embeddings capture the semantic meaning and context of the text, not just word frequency.

Which Evaluation metrics have you used?

Evaluation metric 1	Accuracy
Evaluation metric 2	F1-Score Macro
Evaluation metric 3	F1-Score Weighted
Evaluation metric 4	Classification Report and Confusion Matrix

Add the rows according to the number of evaluation metrics used in the project

Why have you selected those metrics?

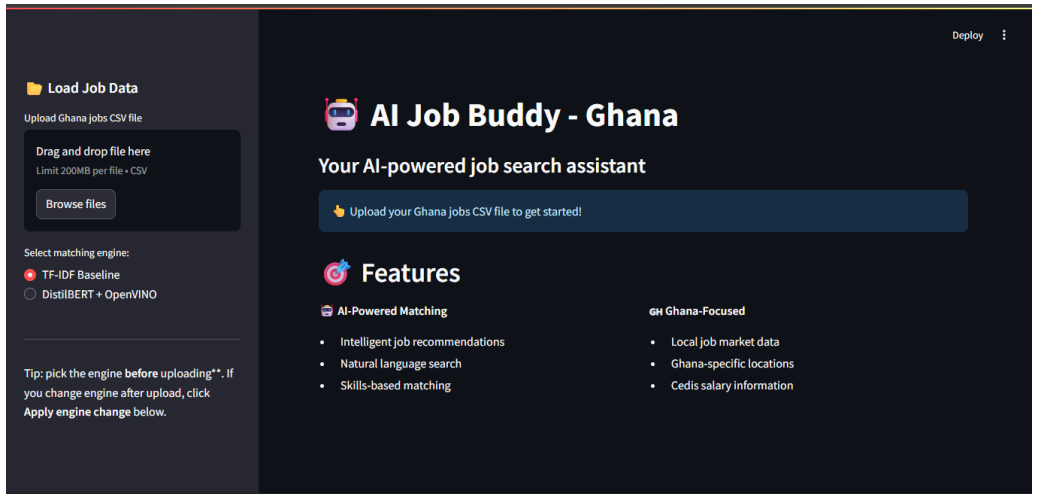
1. Accuracy measures the proportion of correctly classified instances out of the total number of instances. It is a good first-glance metric for overall performance.
2. F1-Score is the harmonic mean of precision and recall, it balances these two metrics to provide a single score that is more robust than accuracy, especially for imbalanced datasets. The F1-Macro metric calculates the F1-score for each class independently and then takes the unweighted average. It treats all categories equally, regardless of their size. The F1-Weighted metric also calculates the F1-score for each class but then takes a weighted average based on the number of instances in each class.
3. The Classification Report is a detailed text summary that provides the precision, recall, and F1-score for each individual class.
4. A Confusion Matrix is a visual table that summarizes the performance of a classification model. The rows represent the actual classes, and the columns represent the predicted classes. This metric provides more insight into a model's errors.

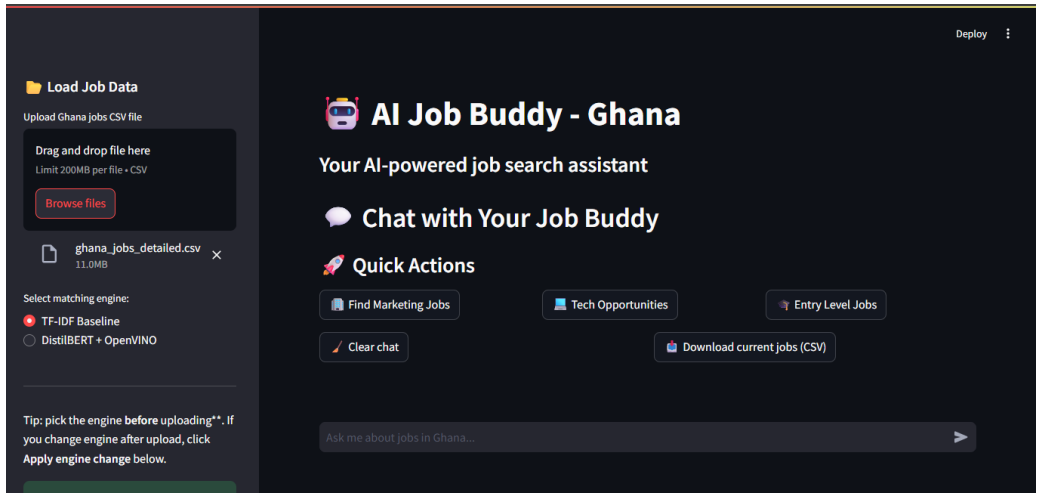
Comparison of the performance of ML model

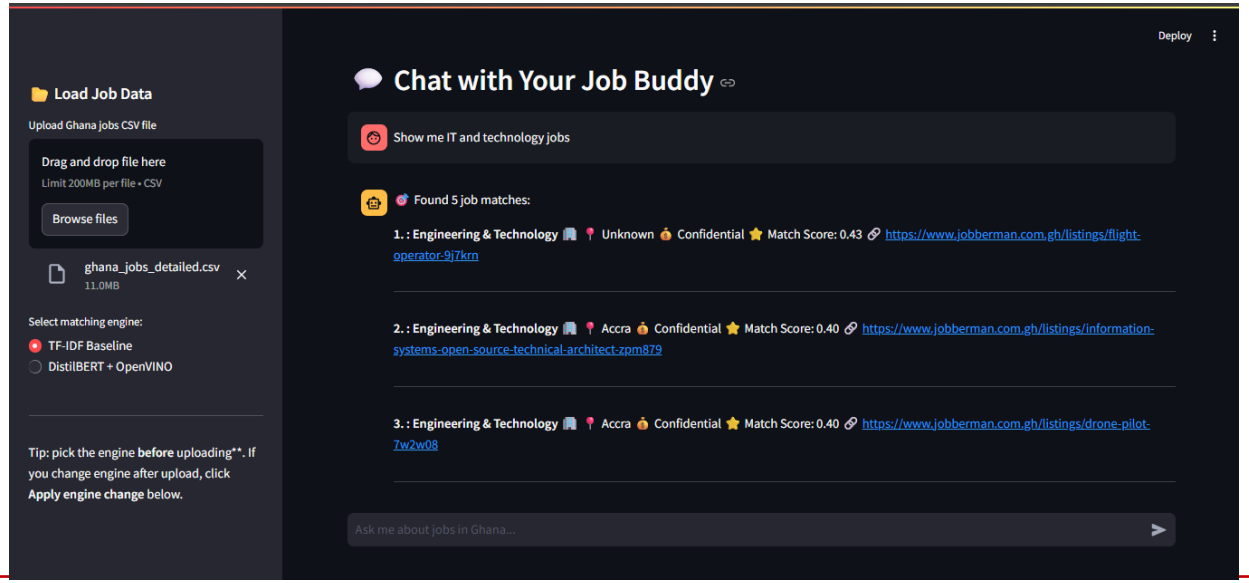
Model	Evaluation metric 1	Evaluation metric 2	Evaluation metric 3
TF-IDF + Cosine Similarity (Baseline)	Accuracy: 58.33%	F1-Macro: 42.90%	F1-Weighted: 59.45%
TF-IDF + Logistic Regression	Accuracy: 63.89%	F1-Macro: 36.32%	F1-Weighted: 64.33%
TF-IDF + RandomForest	Accuracy: 61.11%	F1-Macro: 21.50%	F1-Weighted: 60.57%
DistilBERT (MiniLM) + Logistic Regression	Accuracy: 49.07%	F1-Macro: 37.28%	F1-Weighted: 50.77%

Deployment

Paste the screenshots of the deployed app with all its features here







Paste the images/screenshots here