Azure AI Resume Screener

Nur Izzati Iliyes

May 16th, 2022

CODE; WITHOUT BARRIERS
HACKATHON

CIMB

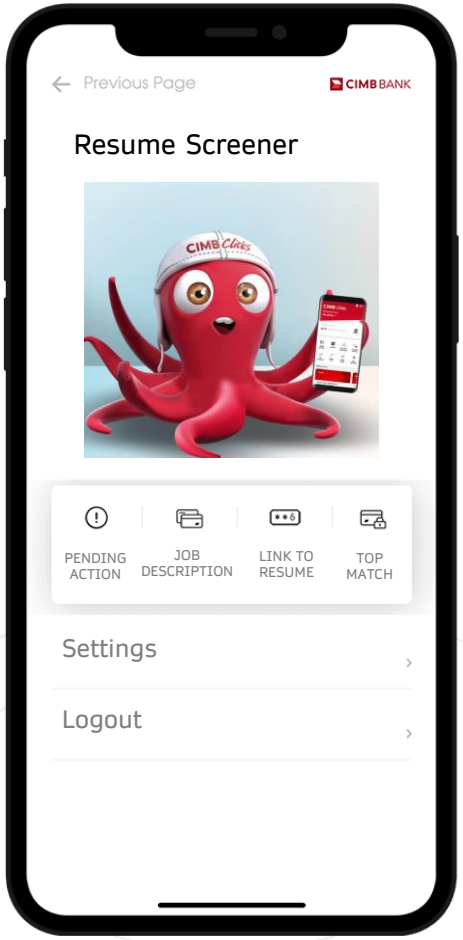
Problem Statement

Recruiter receives huge amounts of resumes everyday and the needs of hiring right talent is crucial. How do we use machine learning to screen and rank the top matching profile to expedite our screening process

Success Criteria

1. Able to shortlist resume based on the skills set needed
2. Able to rank shortlisted resume based one needs of hiring (i.e: years of experience, domain)

The Approach



Front End

PowerApps as UI/UX

- Inputs the Job Description and link to resume repository on Azure blob storage
- Outputs the Top Match Applicants

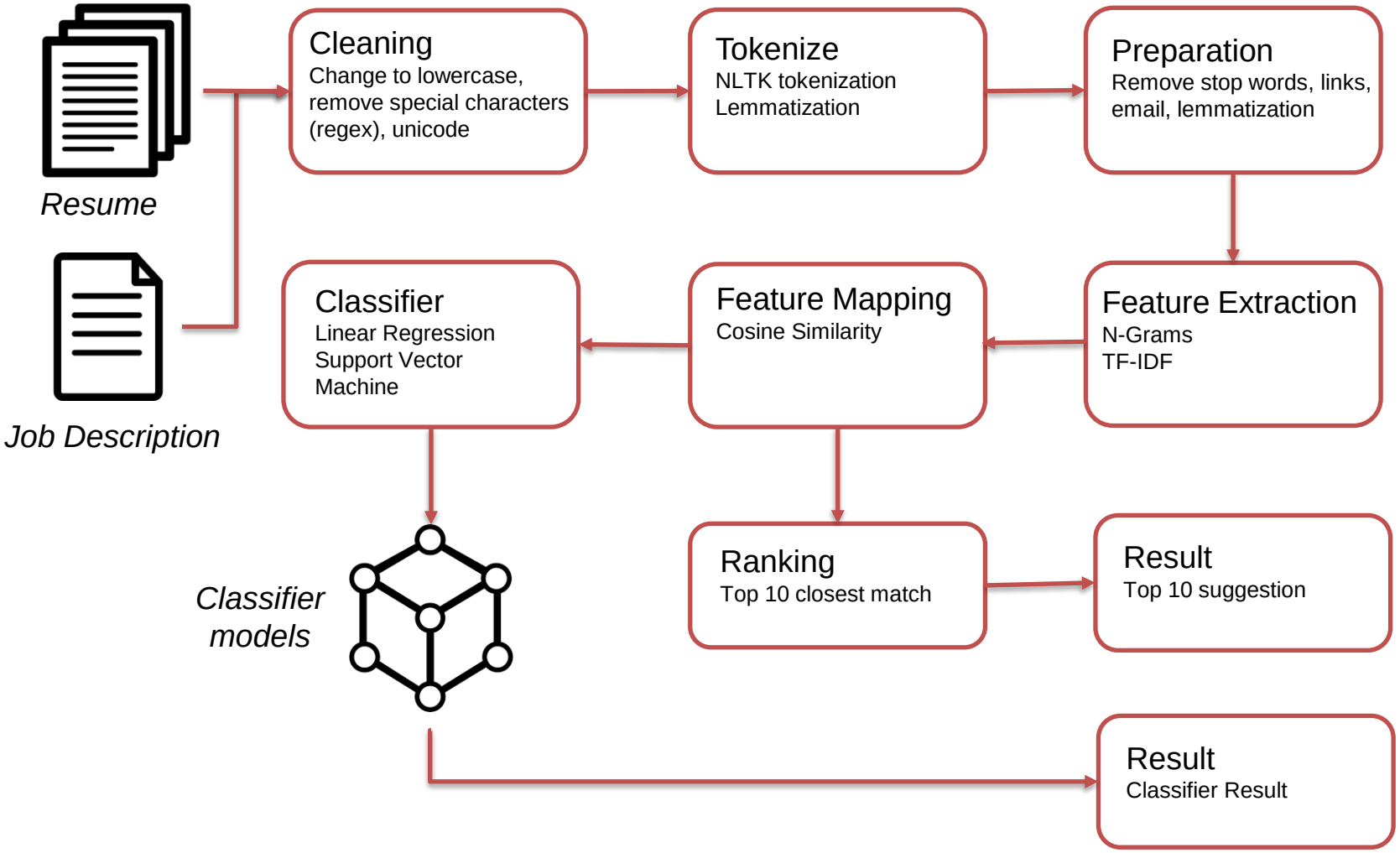
Back End

- Azure blob storage to store JD and resumes
- Azure AI Form Recognizer Python SDK
- Azure Language Cognitive – Key Phrases Extraction & Entity
- Machine Learning Studio to design pipelines for development and deployment to production

CODE WITHOUT BARRIERS HACKATHON

CIMB

The Machine Learning Approach



Challenges of Solving the Hackathon

Data Cleaning

AI Form Recognizer Python SDK does not split words correctly. Need further cleaning to be done before pre-processing.

Training Model Error

Error noted when training the model using ML Designer.

Job Description Template to be standardized

Different templates will have different bounding boxes when Azure does OCR.

Configuring ML Designer Pipeline

Despite being a low/no code application, knowledge up to a certain extent is required to make the overall pipeline work.

Linking Together Output of an AI solution to Another

Putting together an end-to-end AI solution in Azure requires a variety of skills.

Job Description Too Lengthy

A short and precise Job Requirement would make the recommendation system more efficient in screening applicants.

Moving Forward

PRICING

-Cognitive Services

-Compute Instances

-Resource Group (Private or Shared)

SECURITY

-API Key credentials

-Azure RBAC

-Network Security & Containment

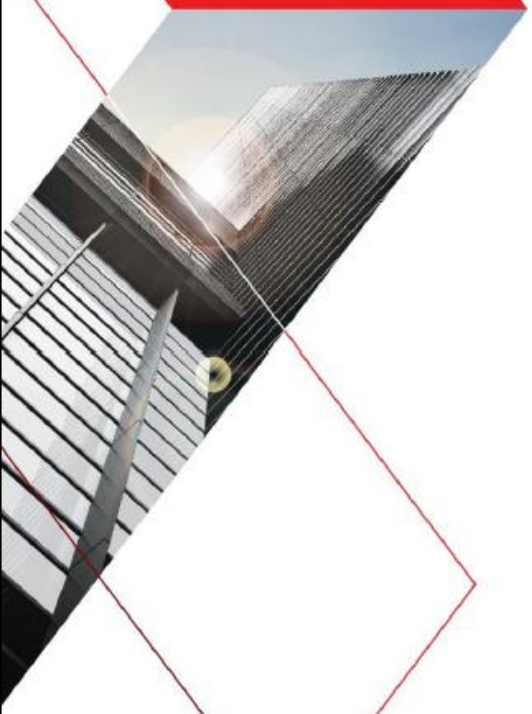
-Storage & Information Protection

SCALABITLITY

-Compute Clusters & Nodes Expansion

-Docker/ Kubernetes deployment

-Azure Storage (ADLS Gen2)



Demo

CODE; WITHOUT BARRIERS
HACKATHON

CIMB

Appendix – Reference

A Machine Learning approach for automation of Resume Recommendation System by Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia (ICCIDS 2019)

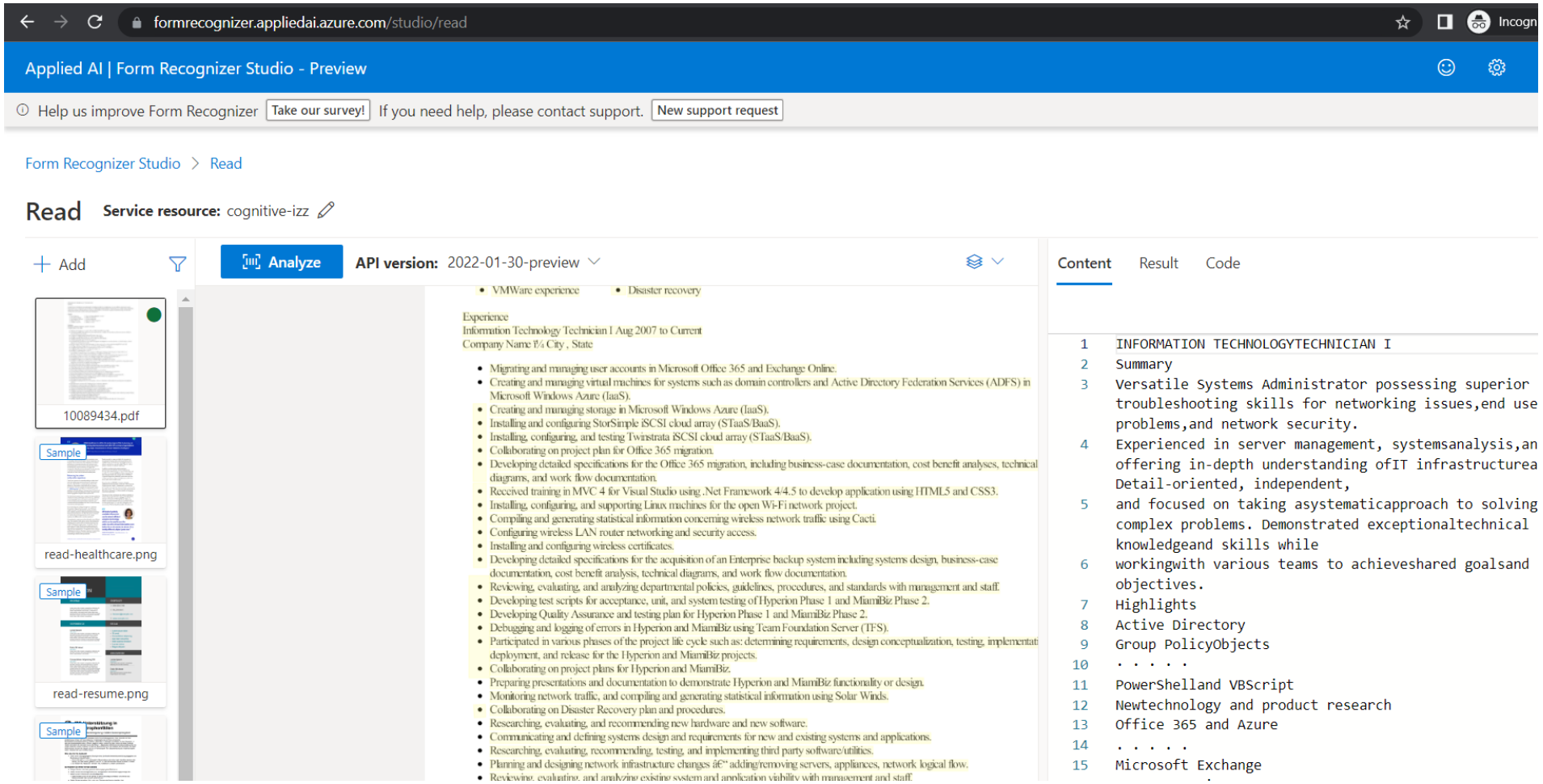
https://www.sciencedirect.com/science/article/pii/S187705092030750X

Microsoft Azure Extract N-Gram Features from Text Analytics

Microsoft Azure Pricing Calculator

https://azure.microsoft.com/en-us/pricing/calculator

Appendix – Form Recognizer Studio



Applied AI | Form Recognizer Studio

CODE WITHOUT BARRIERS HACKATHON

CIMB

Appendix – Form Recognizer Python SDK

read_doc.py | clean_doc.py | compare_and_score.py | extract_keywords_azure.py | extract_keywords_parser.py

read_doc.py > ...

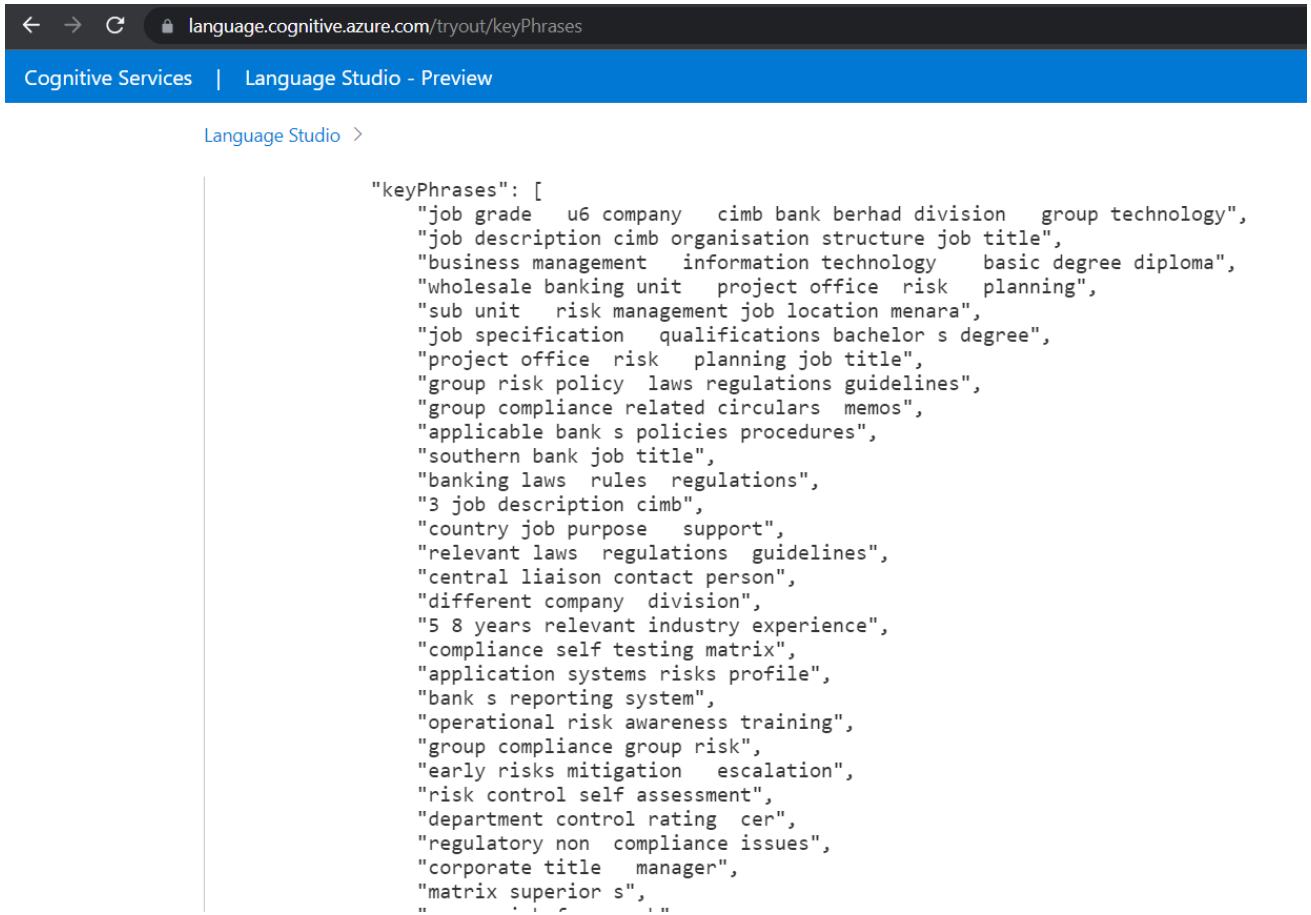
```python
from io import BytesIO
import os
import pandas as pd
from azure.core.exceptions import ResourceNotFoundError
from azure.ai.formrecognizer import FormRecognizerClient, DocumentAnalysisClient
from azure.core.credentials import AzureKeyCredential

API_KEY = "430626c7ec914b88862d1d7b8289fc54"
ENDPOINT = "https://izz-form-recognizer.cognitiveservices.azure.com/"

doc = "10089434.pdf"
#doc = "GT WB - Job Description_IT Risk Lead (RMP)_v1.1.pdf"

form_recognizer_client = FormRecognizerClient(ENDPOINT, AzureKeyCredential(API_KEY))

with open(doc, "rb") as fd:
    f = fd.read()

poller = form_recognizer_client.begin_recognize_content(form=f)
result = poller.result()
form_pages = poller.result()

page = result[0]
# len(page.tables)

data = vars(page)
#print(page)
```

Form Recognizer Python SDK

Name	Date modified	Type
jd.csv	15-May-22 6:29 AM	Microsoft Excel Co..
jd_combine.csv	15-May-22 6:32 PM	Microsoft Excel Co..
jd_stemmed.csv	15-May-22 6:29 AM	Microsoft Excel Co..
jd_stemmed_c.csv	15-May-22 6:29 AM	Microsoft Excel Co..
r.csv	15-May-22 6:29 AM	Microsoft Excel Co..
r_combine.csv	15-May-22 6:32 PM	Microsoft Excel Co..
r_stemmed.csv	15-May-22 6:30 AM	Microsoft Excel Co..
r_stemmed_c.csv	15-May-22 6:30 AM	Microsoft Excel Co..
trained model.csv	15-May-22 6:29 AM	Microsoft Excel Co..

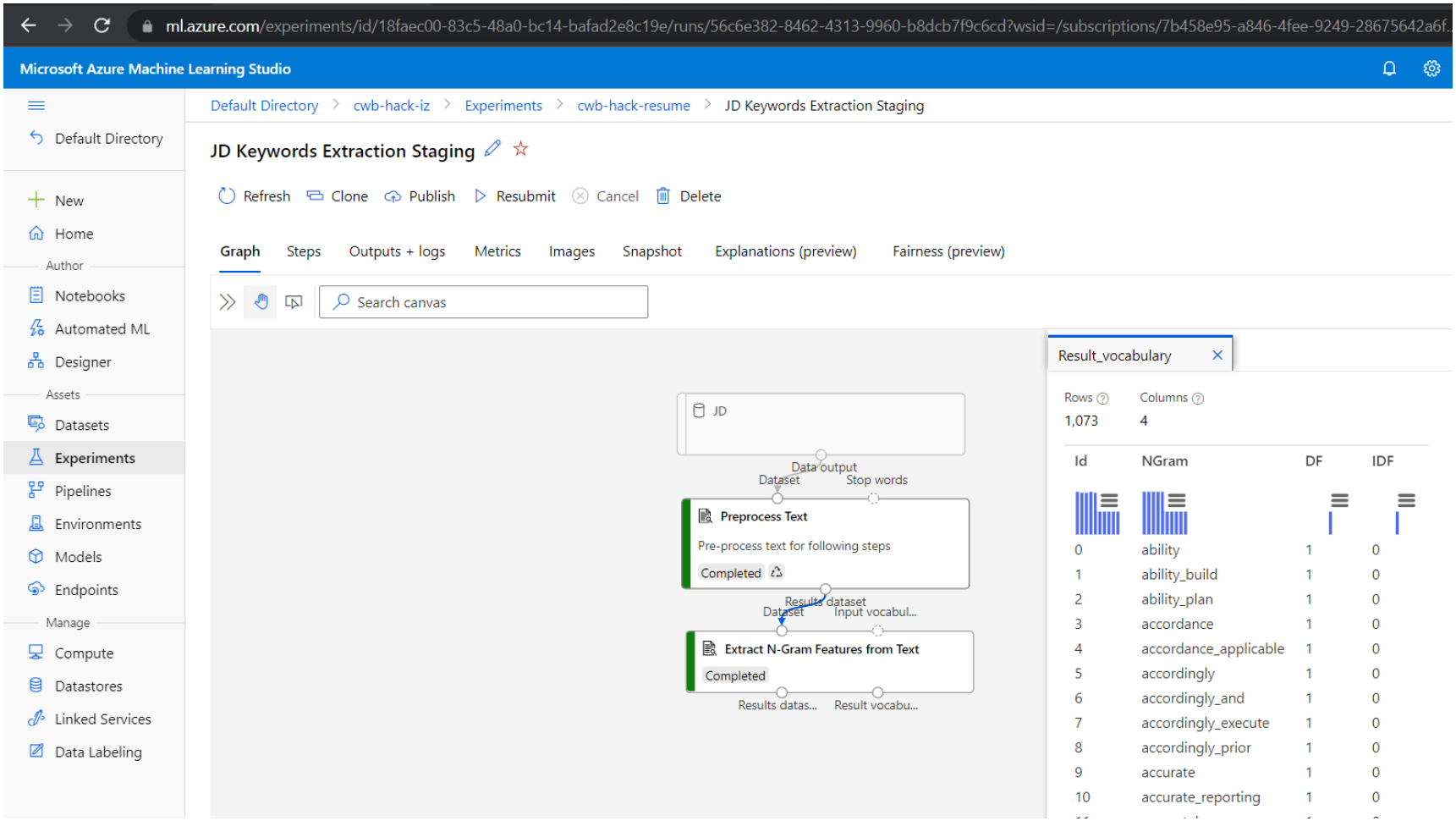
Output Files

Appendix – Language Studio



Cognitive Services | Language Studio

Appendix – Machine Learning Studio

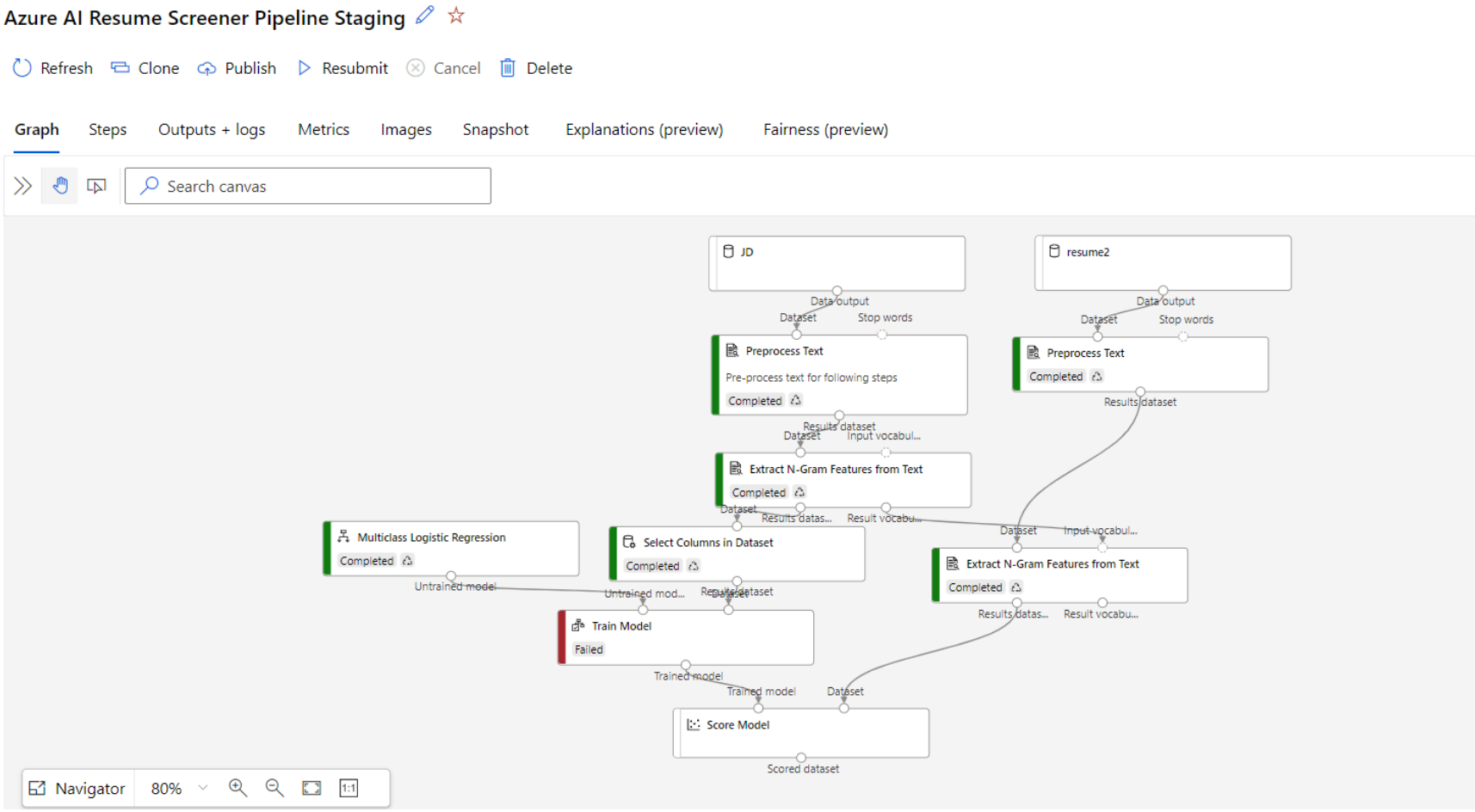


Machine Learning Studio | Keyword Extraction

CODE WITHOUT BARRIERS HACKATHON

CIMB

Appendix – Staging Pipeline

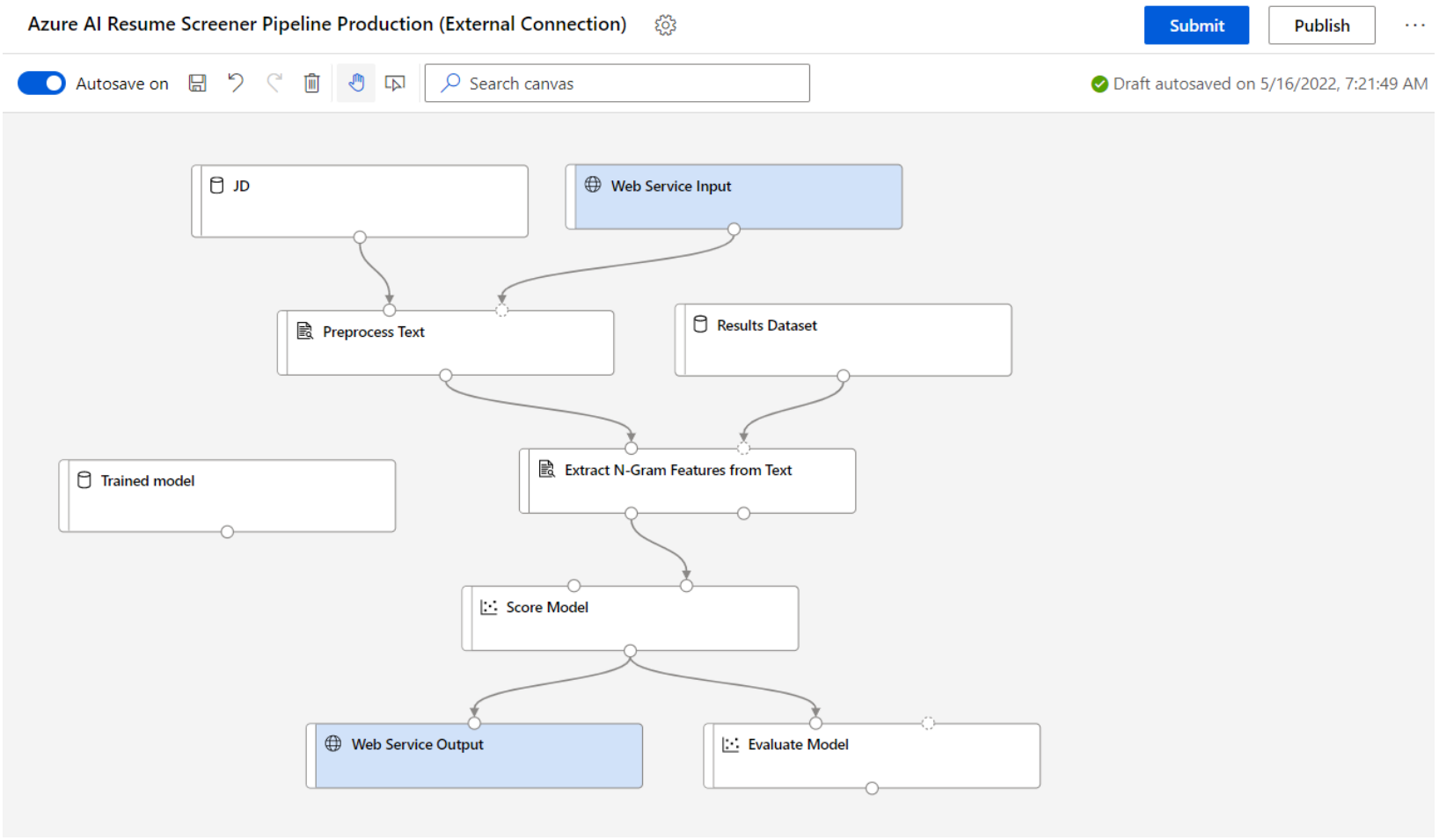


Machine Learning Studio | Development Pipeline

CODE WITHOUT BARRIERS HACKATHON

CIMB

Appendix – Production Pipeline



Machine Learning Studio | Production Pipeline

CODE WITHOUT BARRIERS HACKATHON

CIMB