



Azure AI Resume Screener

Nur Izzati Iliyes

May 16th, 2022

CODE: **WITHOUT BARRIERS**
HACKATHON



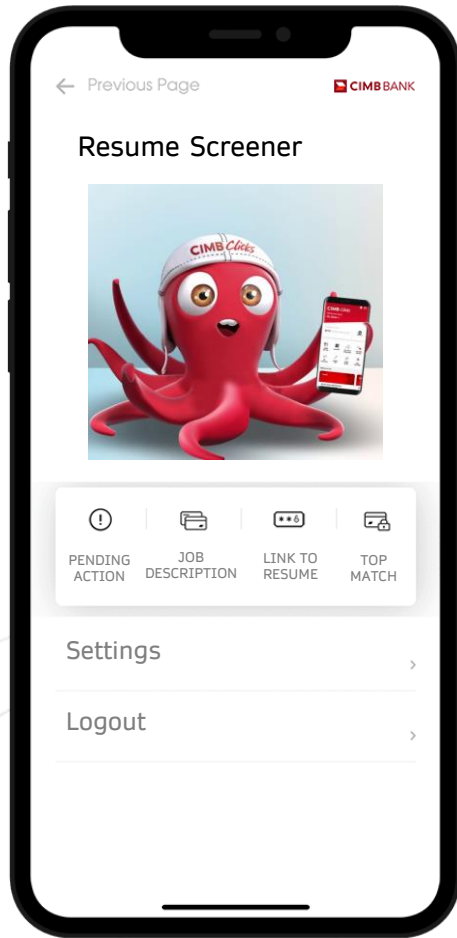
Problem Statement

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 on needs of hiring (i.e: years of experience, domain)

The Approach



Front End

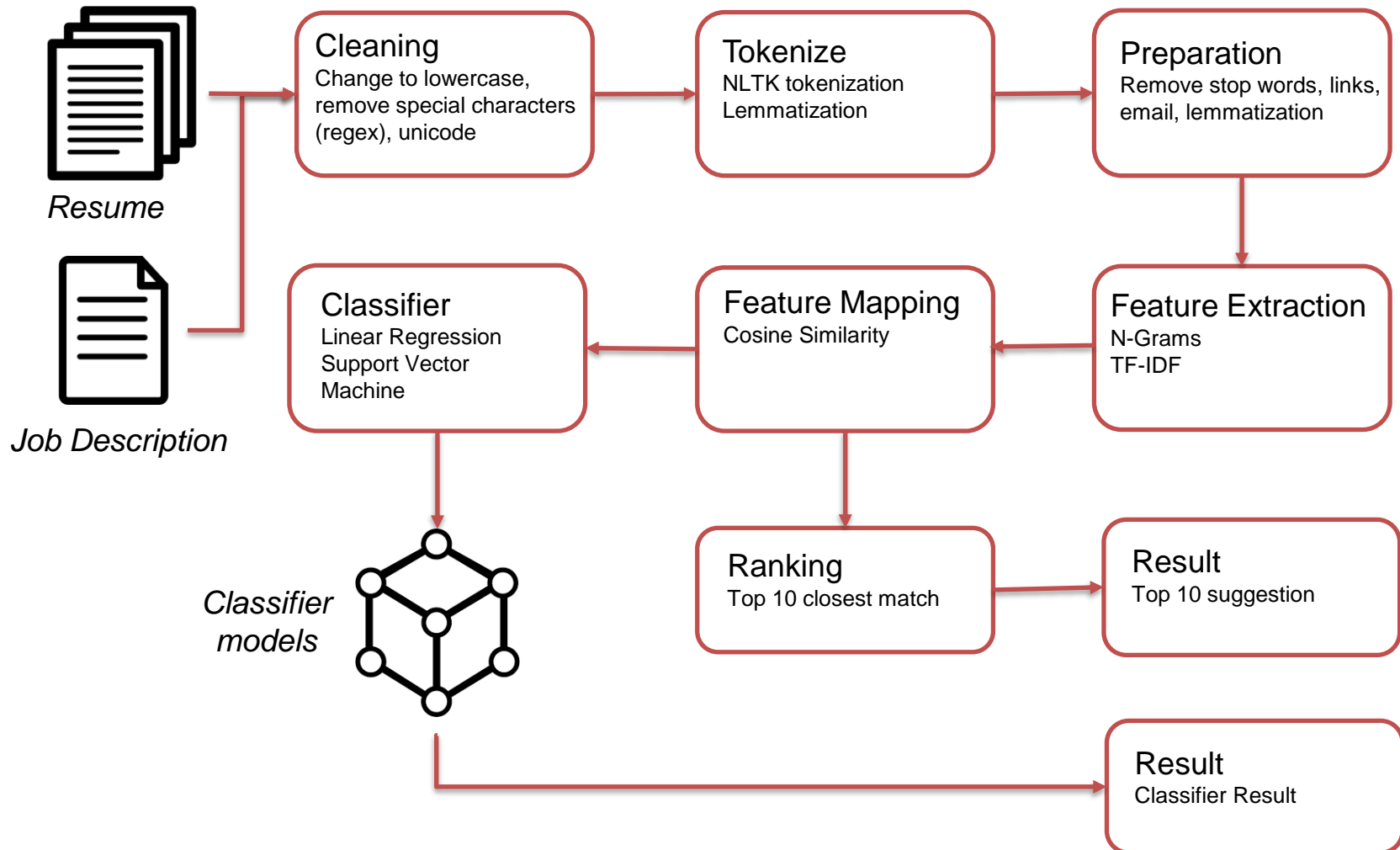
PowerApps as UI/UX

- 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

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

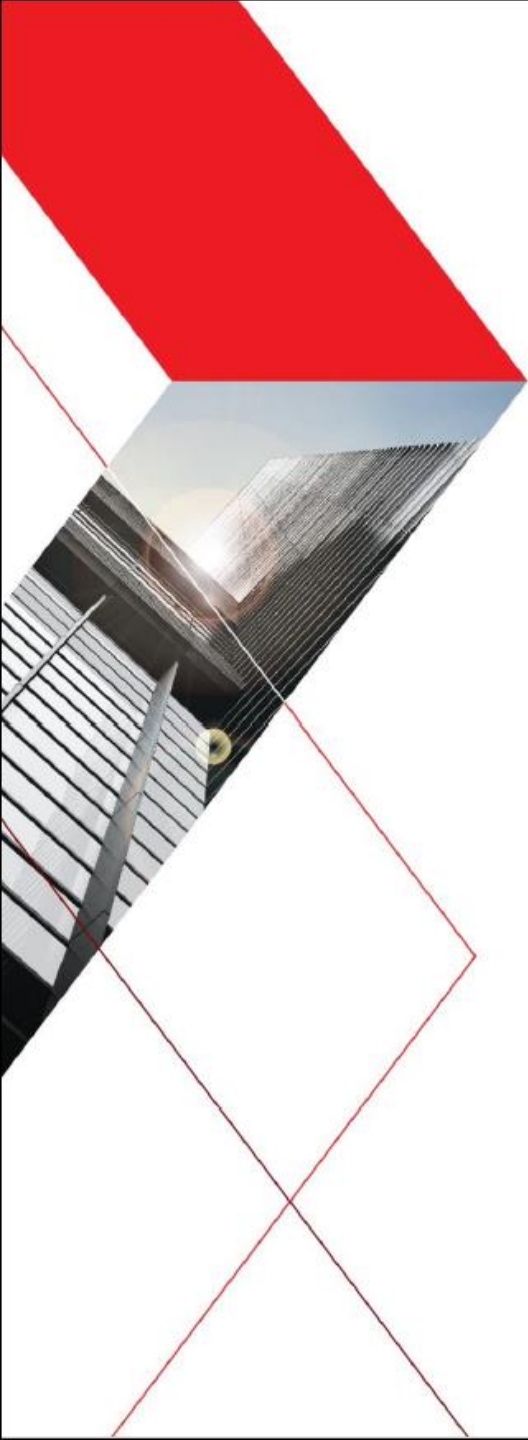
- Cognitive Services
- Compute Instances
- Resource Group (Private or Shared)

SECURITY

- API Key credentials
- Azure RBAC
- Network Security & Containment
- Storage & Information Protection

SCALABILITY

- Compute Clusters & Nodes Expansion
- Docker/ Kubernetes deployment
- Azure Storage (ADLS Gen2)



Demo

CODE: **WITHOUT BARRIERS**
HACKATHON



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

← → ↺ formrecognizer.appliedai.azure.com/studio/read ☆ Incogn

Applied AI | Form Recognizer Studio - Preview

○ Help us improve Form Recognizer [Take our survey!](#) If you need help, please contact support. [New support request](#)

Form Recognizer Studio > Read

Read **Service resource:** cognitive-izz

+ Add

10089434.pdf

Sample

read-healthcare.png

Sample

read-resume.png

Sample

Analyze API version: 2022-01-30-preview

- VMWare experience
- Disaster recovery

Experience
Information Technology Technician I Aug 2007 to Current
Company Name % City , State

- Migrating and managing user accounts in Microsoft Office 365 and Exchange Online.
- Creating and managing virtual machines for systems such as domain controllers and Active Directory Federation Services (ADFS) in Microsoft Windows Azure (IaaS).
- Creating and managing storage in Microsoft Windows Azure (IaaS).
- Installing and configuring StorSimple iSCSI cloud array (STaaS/BaaS).
- Installing, configuring, and testing Twinstrata iSCSI cloud array (STaaS/BaaS).
- Collaborating on project plan for Office 365 migration.
- Developing detailed specifications for the Office 365 migration, including business-case documentation, cost benefit analyses, technical diagrams, and work flow documentation.
- Received training in MVC 4 for Visual Studio using .Net Framework 4/4.5 to develop application using HTML5 and CSS3.
- Installing, configuring, and supporting Linux machines for the open Wi-Fi network project.
- Compiling and generating statistical information concerning wireless network traffic using Cacti.
- Configuring wireless LAN router networking and security access.
- Installing and configuring wireless certificates.
- Developing detailed specifications for the acquisition of an Enterprise backup system including systems design, business-case documentation, cost benefit analysis, technical diagrams, and work flow documentation.
- Reviewing, evaluating, and analyzing departmental policies, guidelines, procedures, and standards with management and staff.
- Developing test scripts for acceptance, unit, and system testing of Hyperion Phase 1 and MiamiBiz Phase 2.
- Developing Quality Assurance and testing plan for Hyperion Phase 1 and MiamiBiz Phase 2.
- Debugging and logging of errors in Hyperion and MiamiBiz using Team Foundation Server (TFS).
- Participated in various phases of the project life cycle such as: determining requirements, design conceptualization, testing, implementation deployment, and release for the Hyperion and MiamiBiz projects.
- Collaborating on project plans for Hyperion and MiamiBiz.
- Preparing presentations and documentation to demonstrate Hyperion and MiamiBiz functionality or design.
- Monitoring network traffic, and compiling and generating statistical information using Solar Winds.
- Collaborating on Disaster Recovery plan and procedures.
- Researching, evaluating, and recommending new hardware and new software.
- Communicating and defining systems design and requirements for new and existing systems and applications.
- Researching, evaluating, recommending, testing, and implementing third party software utilities.
- Planning and designing network infrastructure changes ac" adding/removing servers, appliances, network logical flow.
- Reviewing, evaluating, and analyzing existing system and application viability with management and staff.

Content Result Code

1 INFORMATION TECHNOLOGYTECHNICIAN I

2 Summary

3 Versatile Systems Administrator possessing superior troubleshooting skills for networking issues,end use problems,and network security.

4 Experienced in server management, systemsanalysis,an offering in-depth understanding ofIT infrastructurea

5 Detail-oriented, independent,

6 and focused on taking asystematicapproach to solving complex problems. Demonstrated exceptionaltechnical knowledgeand skills while

7 workingwith various teams to achieveshared goalsand objectives.

8 Highlights

9 Active Directory

10 Group PolicyObjects

11

12 PowerShelland VBScript

13 Newtechnology and product research

14 Office 365 and Azure

15

16 Microsoft Exchange

Applied AI | Form Recognizer Studio

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 > ...
1 from io import BytesIO
2 import os
3 import pandas as pd
4 from azure.core.exceptions import ResourceNotFoundError
5 from azure.ai.formrecognizer import FormRecognizerClient, DocumentAnalysisClient
6 from azure.core.credentials import AzureKeyCredential
7
8 API_KEY = "430626c7ec914b88862d1d7b8289fc54"
9 ENDPOINT = "https://izz-form-recognizer.cognitiveservices.azure.com/"
10
11 doc = "10089434.pdf"
12 #doc = "GT WB - Job Description_IT Risk Lead (RMP)_v1.1.pdf"
13
14 form_recognizer_client = FormRecognizerClient(ENDPOINT, AzureKeyCredential(API_KEY))
15
16 with open(doc, "rb") as fd:
17     f = fd.read()
18
19 poller = form_recognizer_client.begin_recognize_content(form=f)
20 result = poller.result()
21 form_pages = poller.result()
22
23 page = result[0]
24 # len(page.tables)
25
26 data = vars(page)
27 #print(page)
28
```

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



Language Studio >

```
"keyPhrases": [
  "job grade u6 company cimb bank berhad division group technology",
  "job description cimb organisation structure job title",
  "business management information technology basic degree diploma",
  "wholesale banking unit project office risk planning",
  "sub unit risk management job location menara",
  "job specification qualifications bachelor s degree",
  "project office risk planning job title",
  "group risk policy laws regulations guidelines",
  "group compliance related circulars memos",
  "applicable bank s policies procedures",
  "southern bank job title",
  "banking laws rules regulations",
  "3 job description cimb",
  "country job purpose support",
  "relevant laws regulations guidelines",
  "central liaison contact person",
  "different company division",
  "5 8 years relevant industry experience",
  "compliance self testing matrix",
  "application systems risks profile",
  "bank s reporting system",
  "operational risk awareness training",
  "group compliance group risk",
  "early risks mitigation escalation",
  "risk control self assessment",
  "department control rating cer",
  "regulatory non compliance issues",
  "corporate title manager",
  "matrix superior s",
  "job title"
]
```

Cognitive Services | Language Studio

Appendix – Machine Learning Studio

ml.azure.com/experiments/id/18faec00-83c5-48a0-bc14-bafad2e8c19e/runs/56c6e382-8462-4313-9960-b8dcb7f9c6cd?wsid=/subscriptions/7b458e95-a846-4fee-9249-28675642a6f...

Microsoft Azure Machine Learning Studio

Default Directory > cwb-hack-iz > Experiments > cwb-hack-resume > JD Keywords Extraction Staging

JD Keywords Extraction Staging

Refresh Clone Publish Resubmit Cancel Delete

Graph Steps Outputs + logs Metrics Images Snapshot Explanations (preview) Fairness (preview)

Search canvas







Result_vocabulary

Id	NGram	DF	IDF
0	ability	1	0
1	ability_build	1	0
2	ability_plan	1	0
3	accordance	1	0
4	accordance_applicable	1	0
5	accordingly	1	0
6	accordingly_and	1	0
7	accordingly_execute	1	0
8	accordingly_prior	1	0
9	accurate	1	0
10	accurate_reporting	1	0

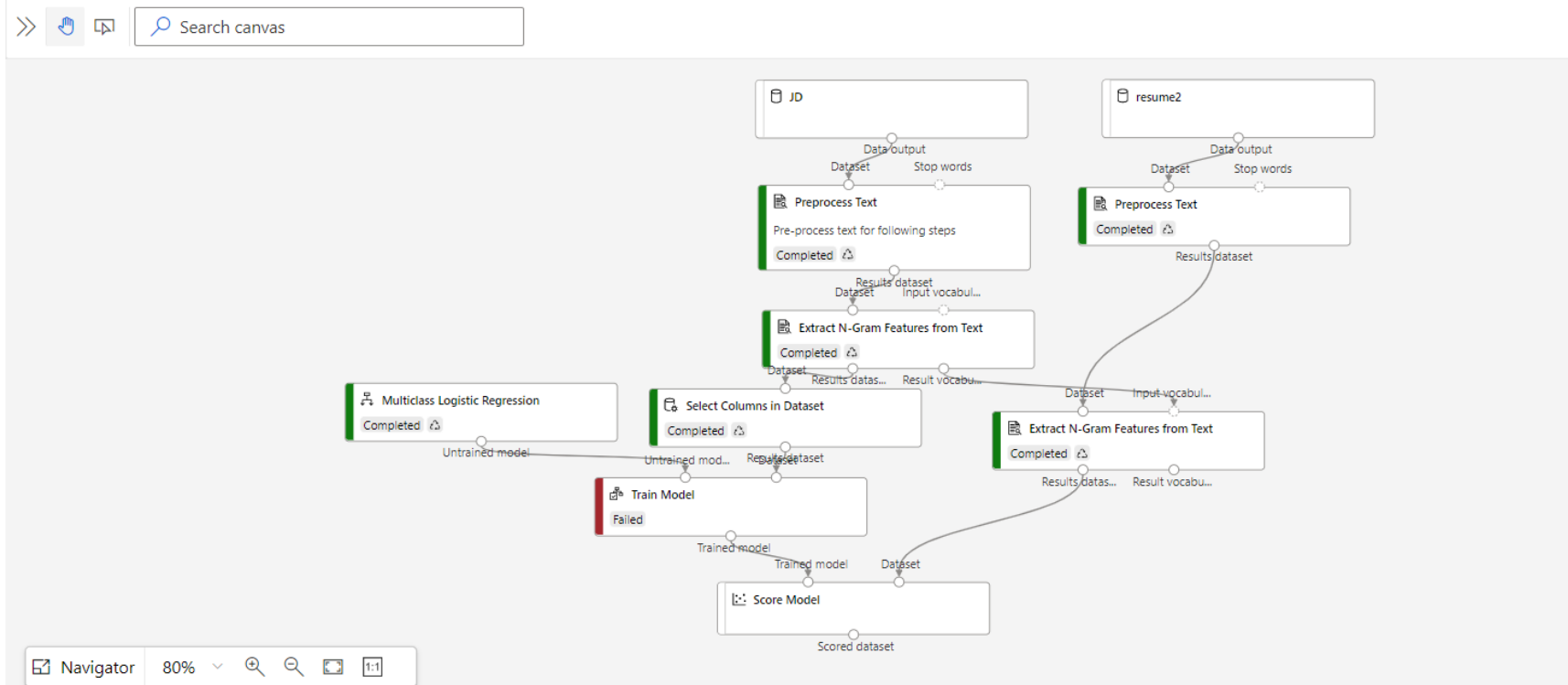
Machine Learning Studio | Keyword Extraction

Appendix – Staging Pipeline

Azure AI Resume Screener Pipeline Staging

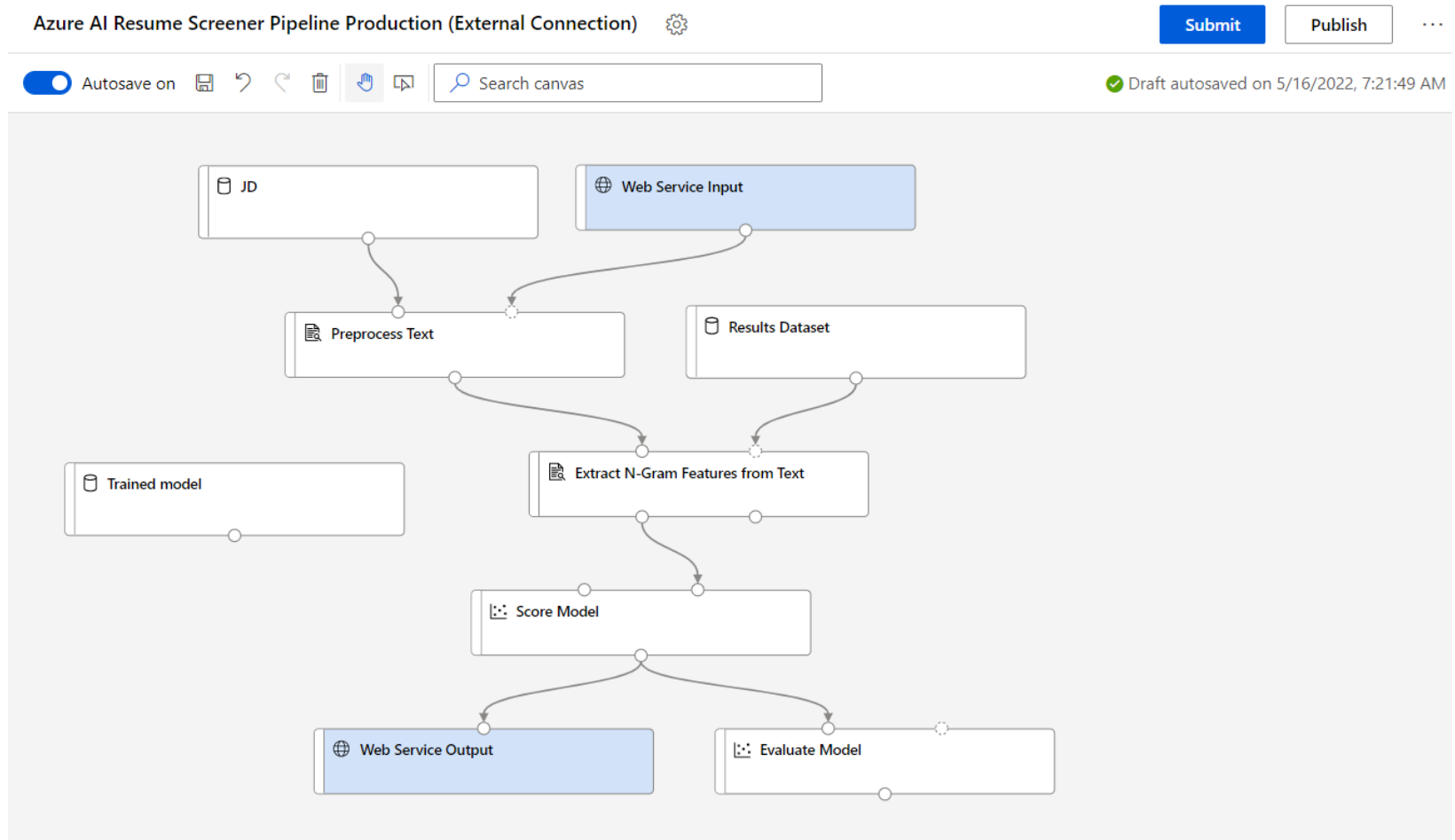
 Refresh  Clone  Publish  Resubmit  Cancel  Delete

Graph Steps Outputs + logs Metrics Images Snapshot Explanations (preview) Fairness (preview)



Machine Learning Studio | Development Pipeline

Appendix – Production Pipeline



Machine Learning Studio | Production Pipeline