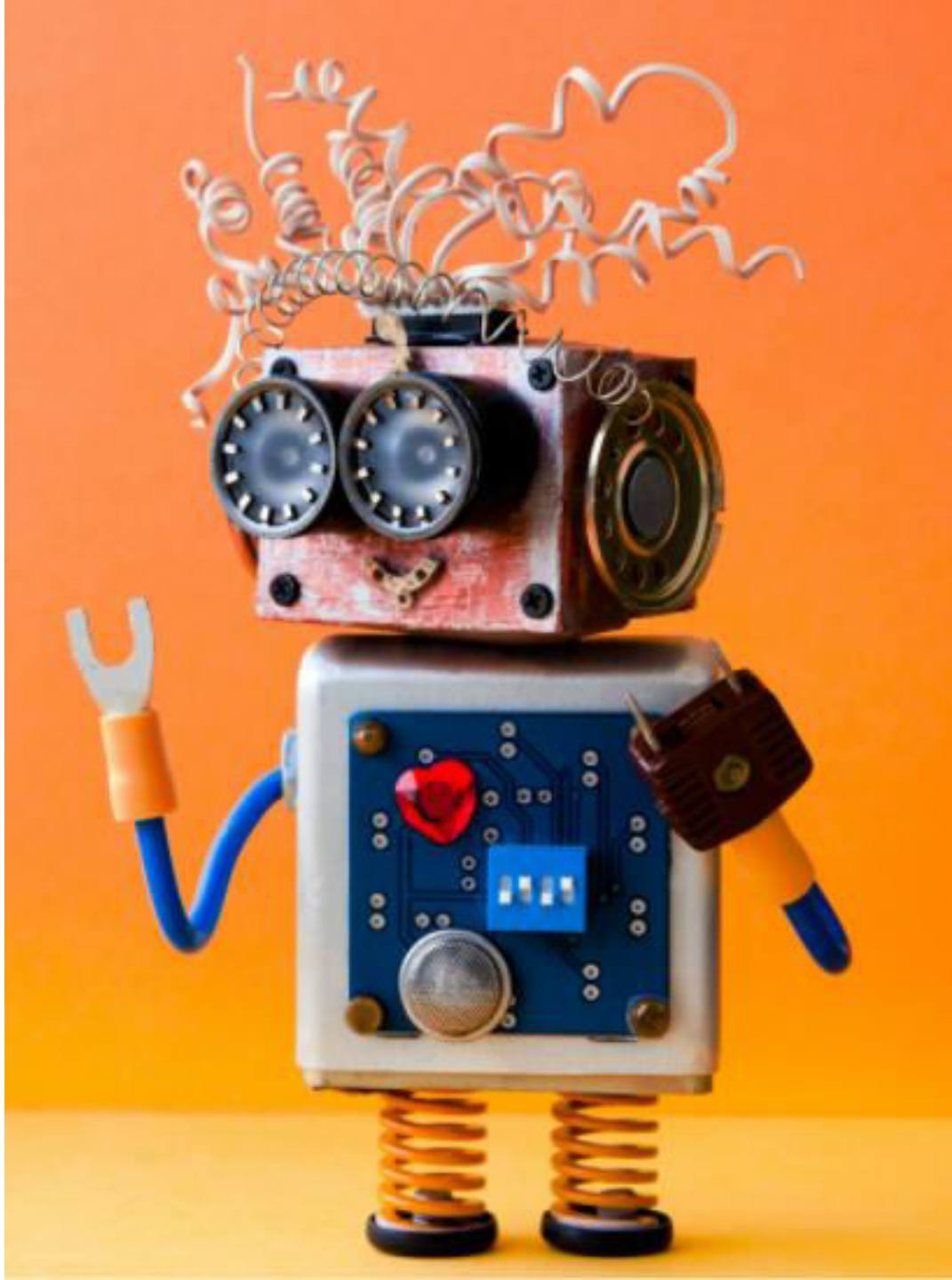
PROJECT CARD

EASY

ADVANCED



PROJECT CARD

GOAL:

- *Build, train, test and deploy a machine learning model to predict chances of university admission into a particular university* given student's profile.

TOOL:

- *AWS SageMaker – Launching a Training job from the Management Console*

PRACTICAL REAL-WORLD APPLICATION:

- *This project can be effectively used by university admission departments to determine top qualifying students.*

DATA:

INPUTS (FEATURES):

- *GRE Scores (out of 340)*
- *TOEFL Scores (out of 120)*
- *University Rating (out of 5)*
- *Statement of Purpose (SOP)*
- *Letter of Recommendation (LOR) Strength (out of 5)*
- *Undergraduate GPA (out of 10)*
- *Research Experience (either 0 or 1)*

OUTPUTS:

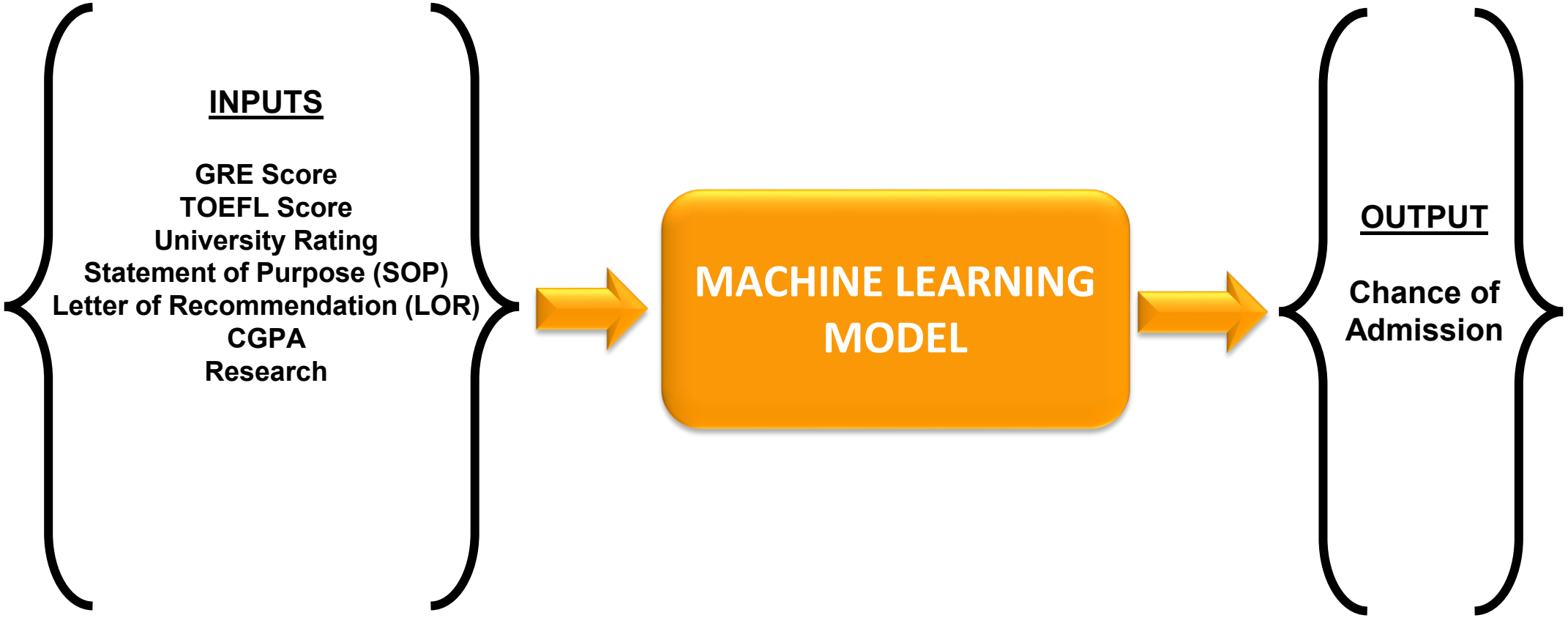
- *Chance of admission (ranging from 0 to 1)*

- Data Source: https://www.kaggle.com/mohansacharya/graduate-admissions
- Photo Credit: https://www.pexels.com/photo/accomplishment-ceremony-education-graduation-267885/

https://www.flickr.com/photos/pasa/6767993805

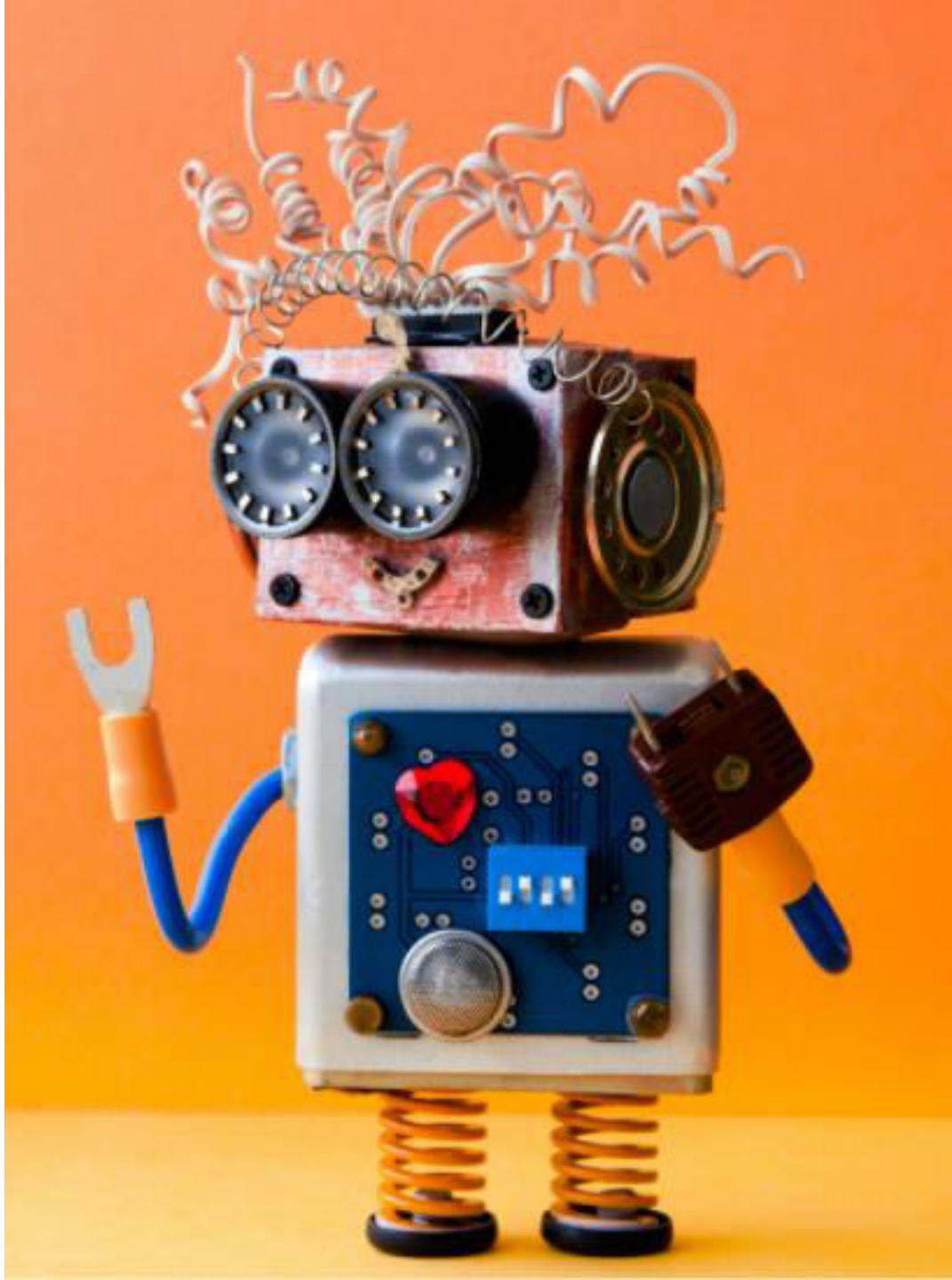
https://www.kaggle.com/ljanjughazyan/cars1

PROJECT OVERVIEW



THE RISE OF MACHINE LEARNING IN HIGHER EDUCATION

EASY ADVANCED



THE RISE OF MACHINE LEARNING IN HIGHER EDUCATION

- Machine Learning and Artificial Intelligence have been transforming higher education in many areas such as:

Marketing and Recruiting

Students Admission and Enrollment

Curriculum and Resources Planning and Forecasting

Pedagogy and Personalized Students Learning Experience

Students Support (AI-Powered Counseling)

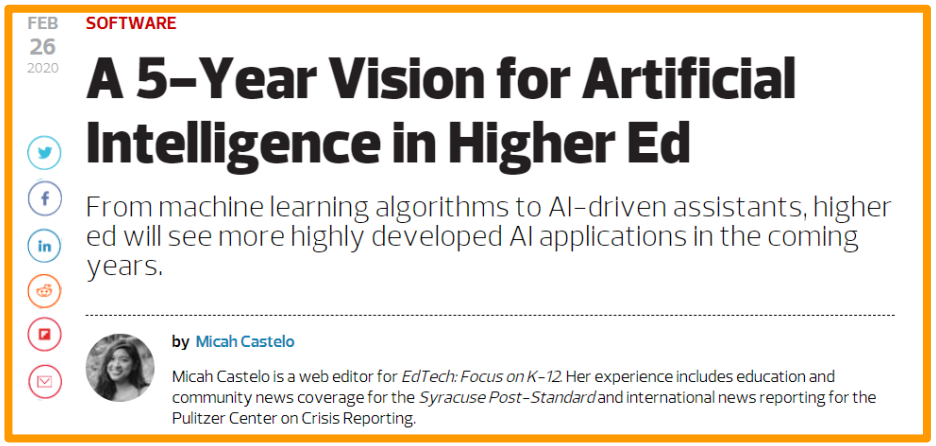
- Reference: *https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions*

READING TIME & QUIZ: THE RISE OF AI IN HIGHER EDUCATION

- Please read the 2 articles below and answer the following quiz.
  - https://edtechmagazine.com/higher/article/2020/02/5-year-vision-artificial-intelligence-higher-ed
  - https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions

15 MINS

10 MINS



FEB 26 2020

SOFTWARE

A 5-Year Vision for Artificial Intelligence in Higher Ed

From machine learning algorithms to AI-driven assistants, higher ed will see more highly developed AI applications in the coming years.

by Micah Castelo

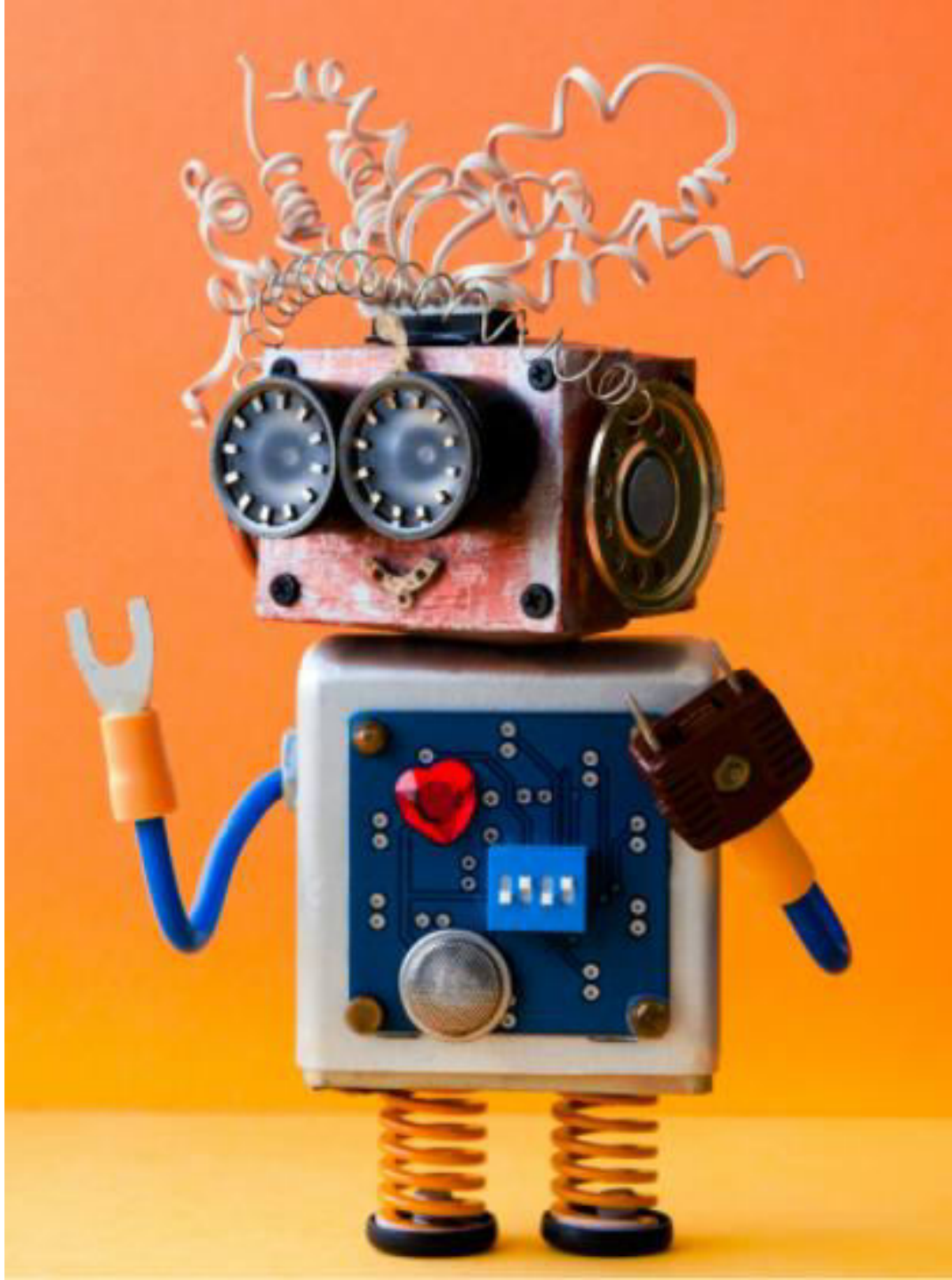
Micah Castelo is a web editor for *EdTech: Focus on K–12*. Her experience includes education and community news coverage for the *Syracuse Post-Standard* and international news reporting for the Pulitzer Center on Crisis Reporting.

Artificial Intelligence in Higher Education: Applications, Promise and Perils, and Ethical Questions

Elana Zeide Monday, August 26, 2019 In Print PDF

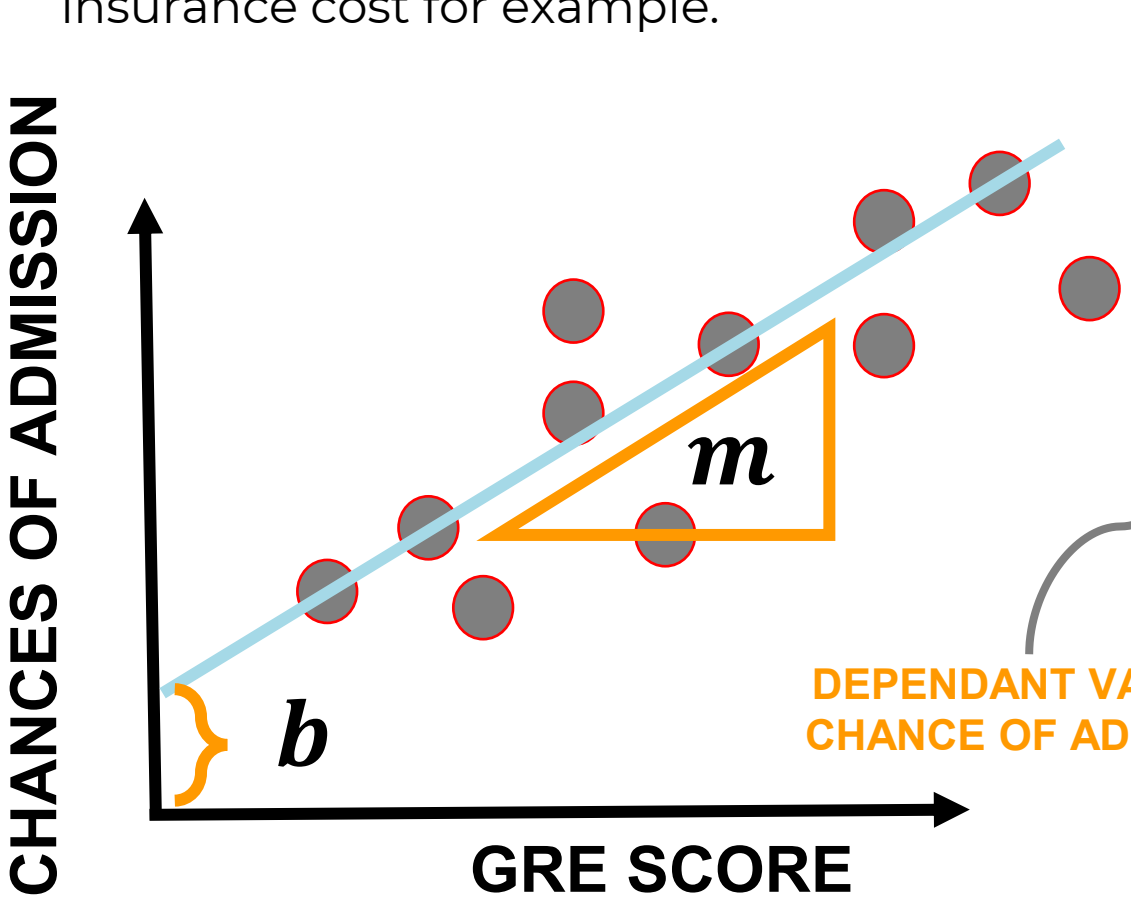
16 min read

MACHINE LEARNING REGRESSION RECAP



SIMPLE LINEAR REGRESSION

- Regression works by predicting value of one variable Y based on another variable X.
- X is called the independent variable and Y is called the dependant variable.
- Goal is to obtain a relationship (model) between two variables only such as age and insurance cost for example.



MODEL! (GOAL)

$$y = b + m * x$$

DEPENDANT VARIABLE
CHANCE OF ADMISSION

INDEPENDENT VARIABLE
GRE SCORE

MULTIPLE LINEAR REGRESSION

- Multiple Linear Regression: examines relationship between more than two variables.
- Recall that Simple Linear regression is a statistical model that examines linear relationship between two variables only.
- Each independent variable has its own corresponding coefficient.

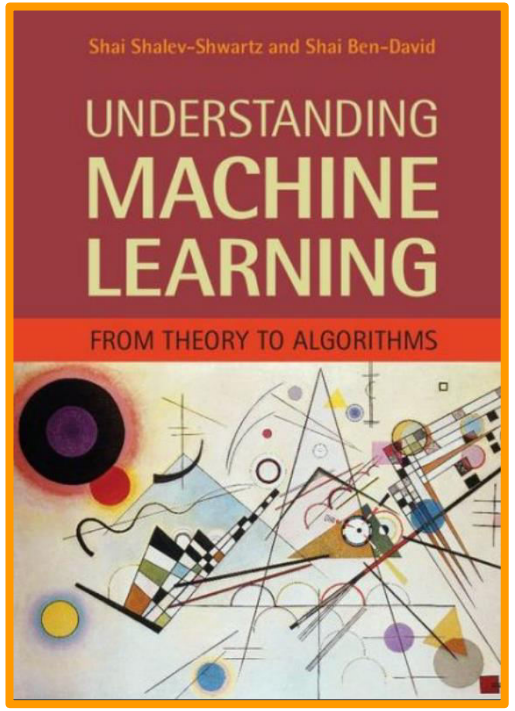
$$y = b_0 + b_1 * x_1 + b_2 * x_2 + .. + b_n x_n$$

CHANCE OF ADMISSION (RANGE 0-1)

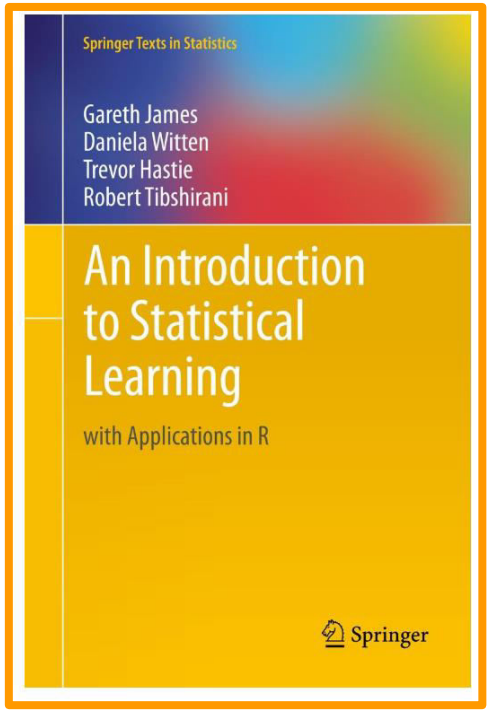
INDEPENDENT VARIABLES (GRE SCORE, GPA, TOEFL SCORES..ETC)

ADDITIONAL READING MATERIAL

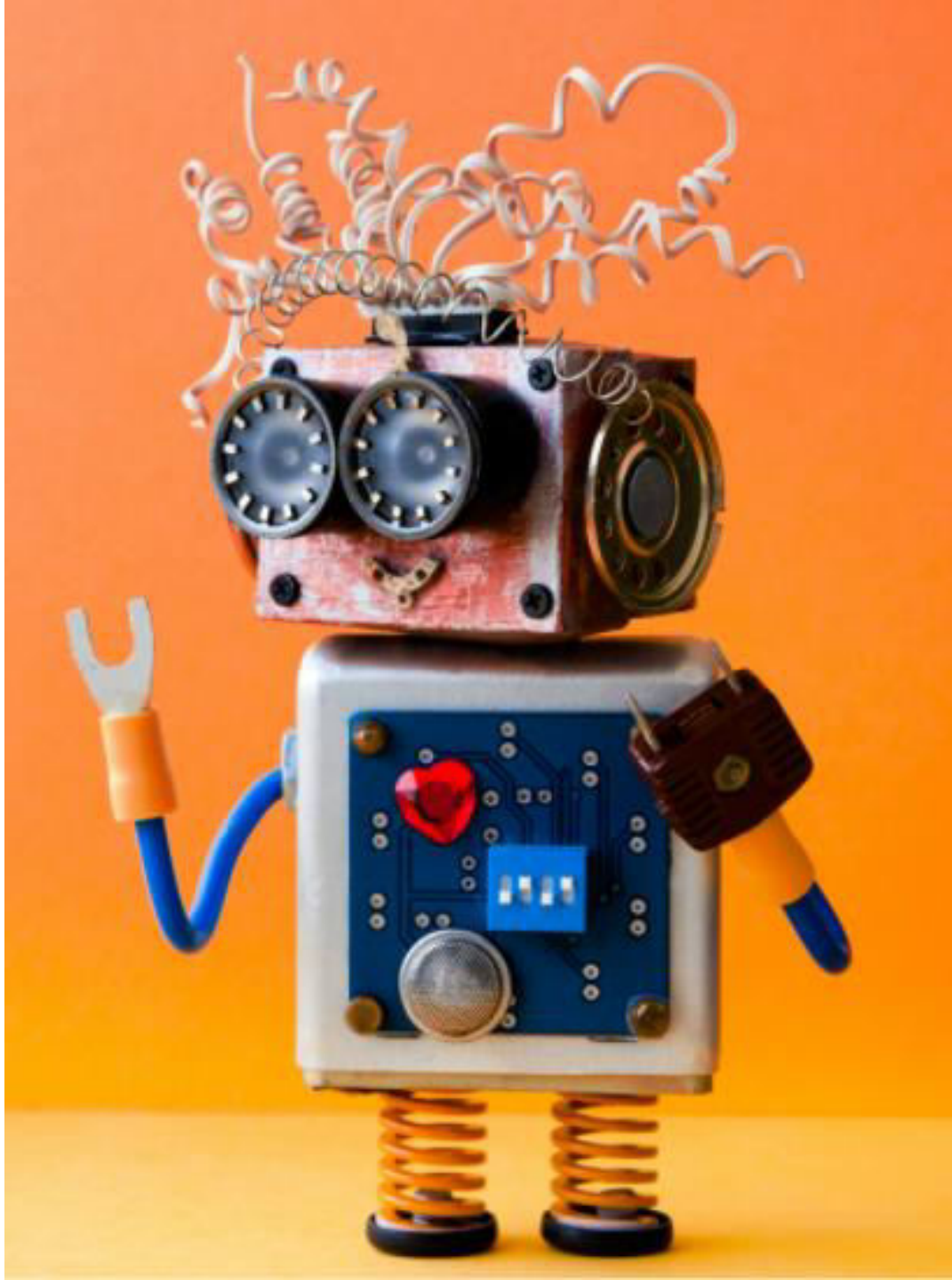
Additional Resources, Page #123:
http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf



Additional Resources, Page #61:
http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf



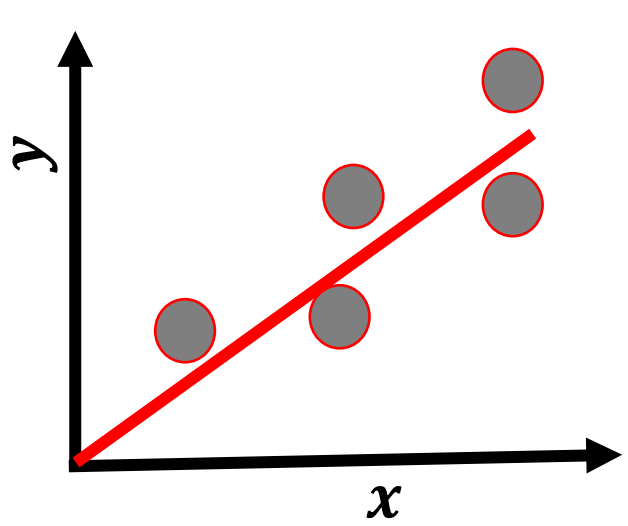
PRACTICE OPPORTUNITY

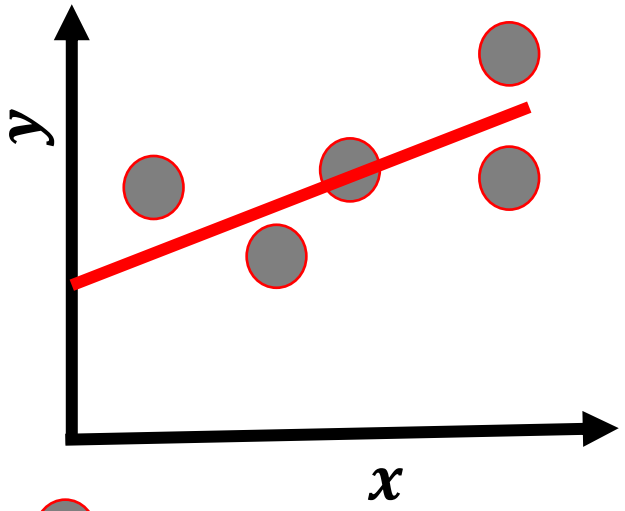


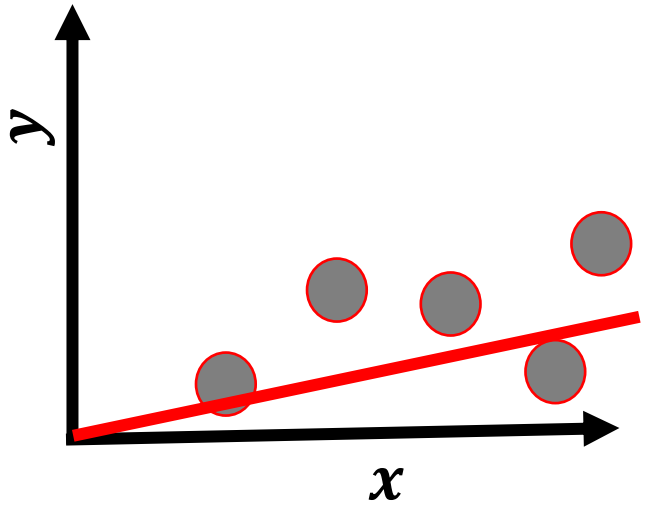
PRACTICE OPPORTUNITY:

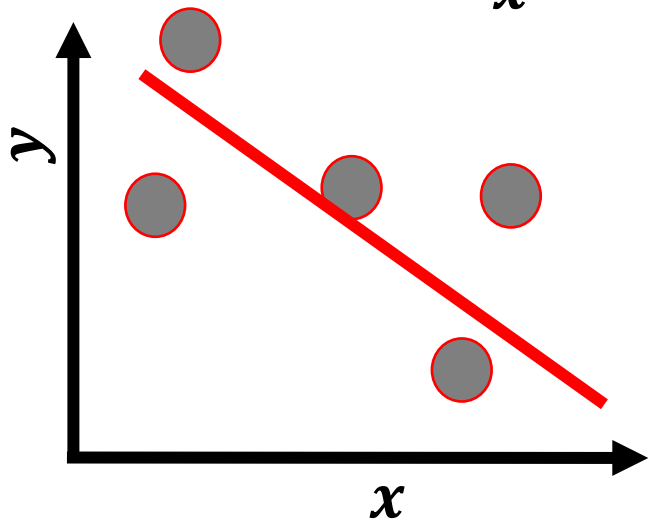
- Match the equations to the figures below and explain why:

$$y = 3 * x + 5$$
$$y = 1 * x$$
$$y = -3 * x + 4$$
$$y = 4 * x$$



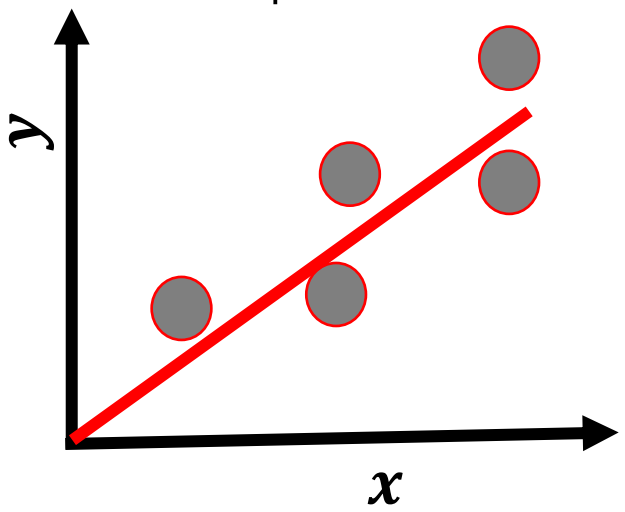




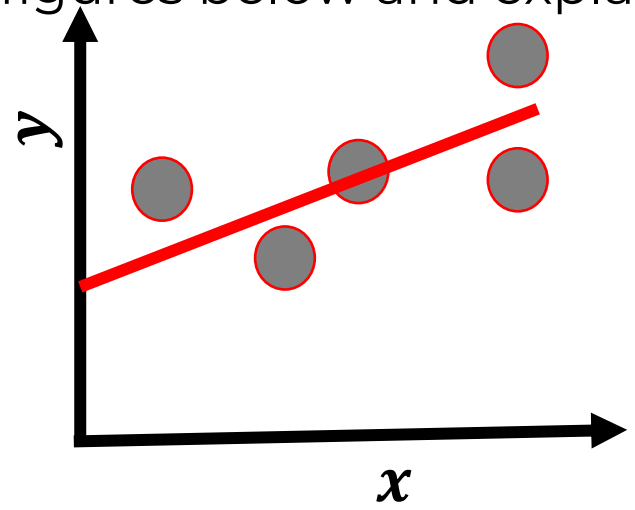


PRACTICE OPPORTUNITY SOLUTION:

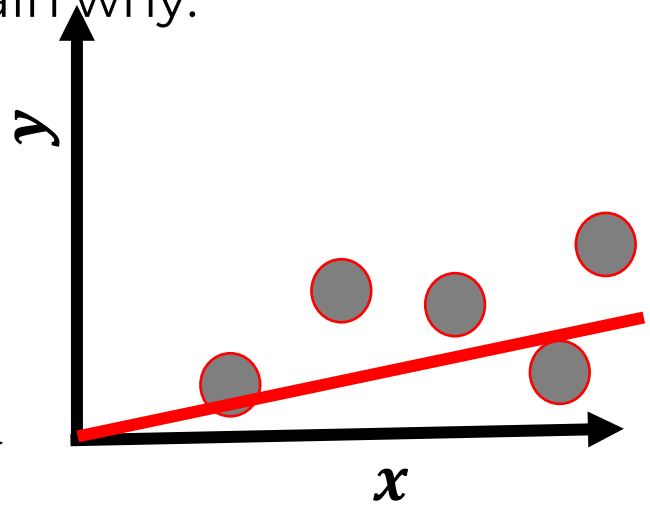
- Match the equations to the figures below and explain why:



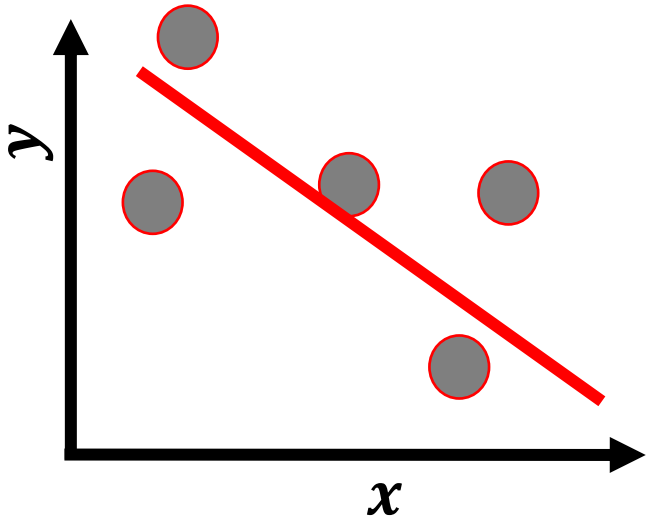
$y = 4 * x$



$y = 3 * x + 5$

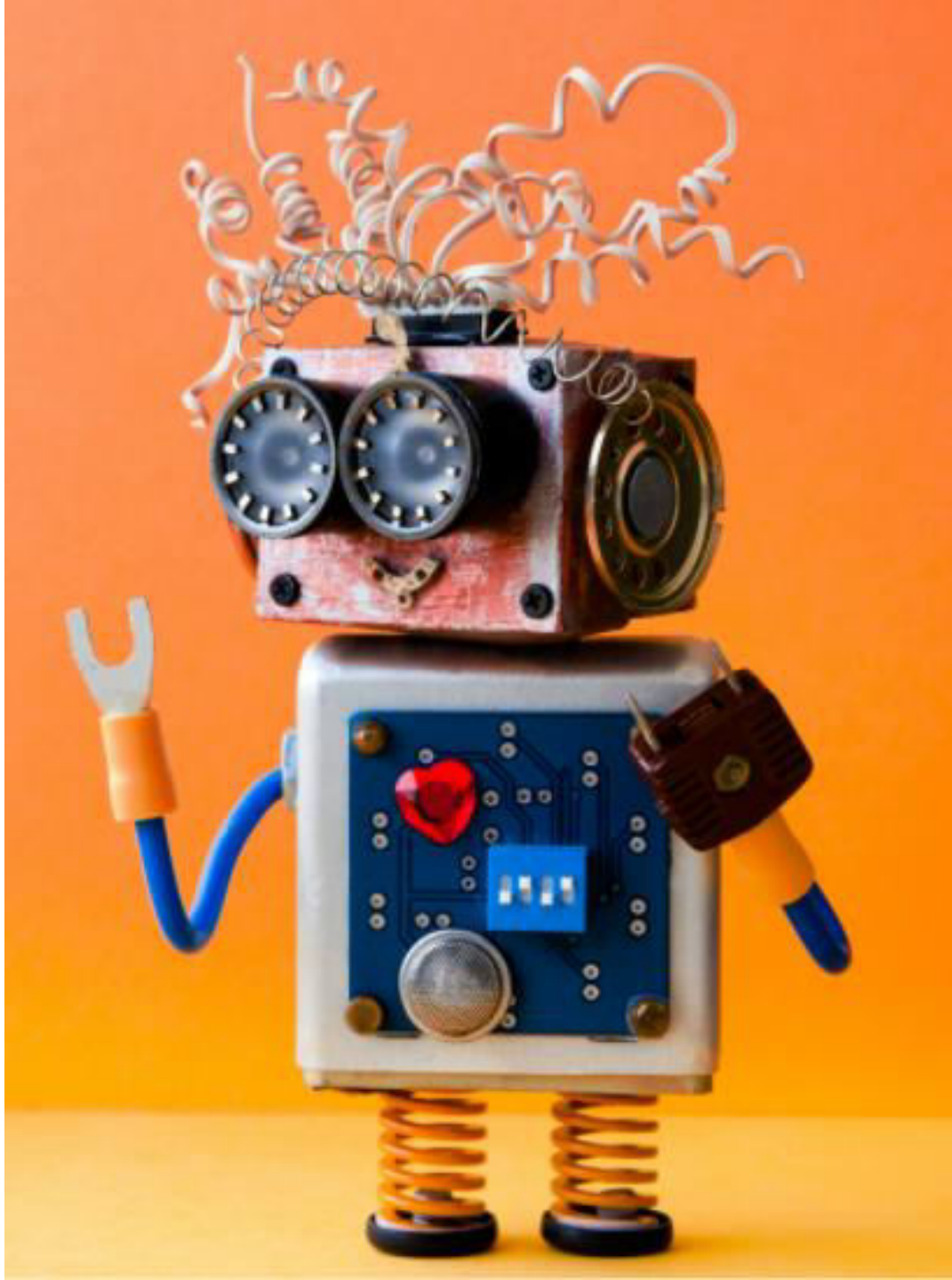


$y = 1 * x$



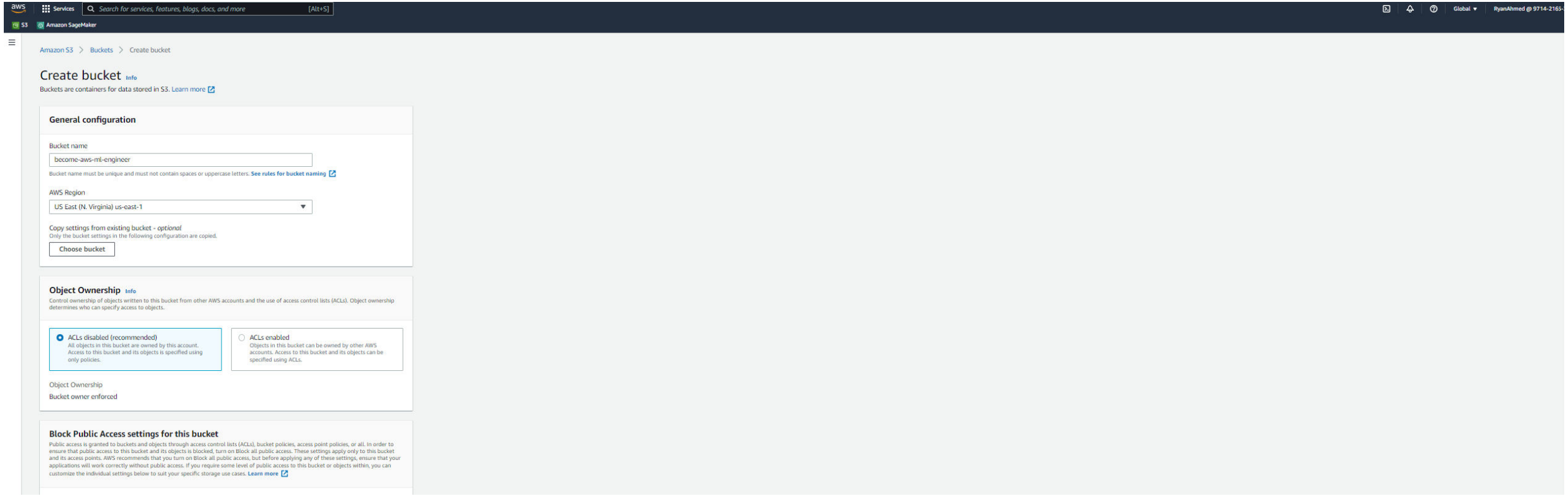
$y = -3 * x + 4$

LAUNCH A TRAINING JOB FROM AWS CONSOLE DEMO PART #1



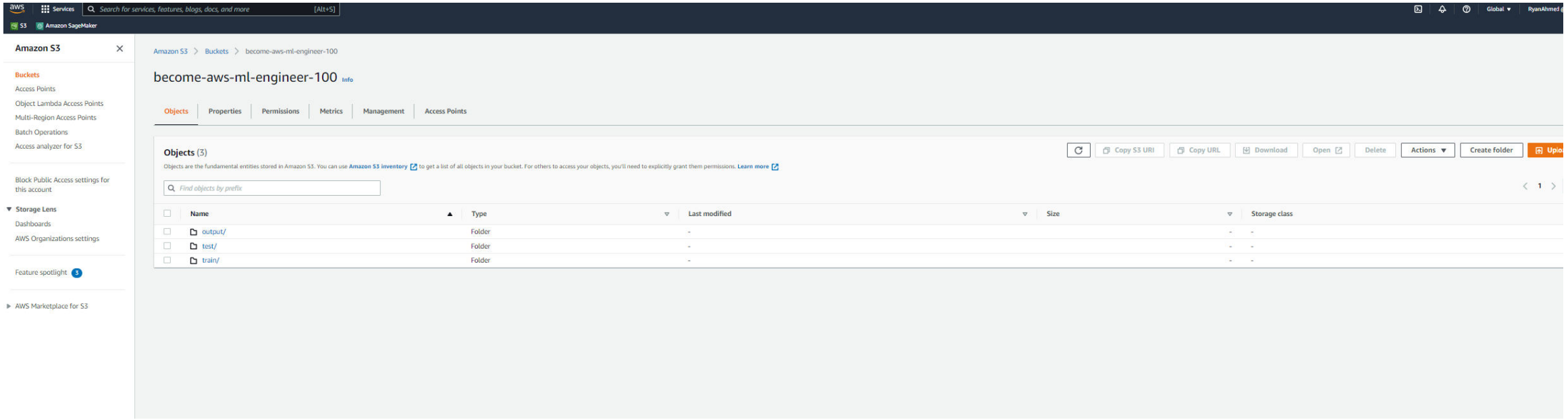
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

CREATE AN S3 BUCKET AND UPLOAD THE TRAINING DATASETS



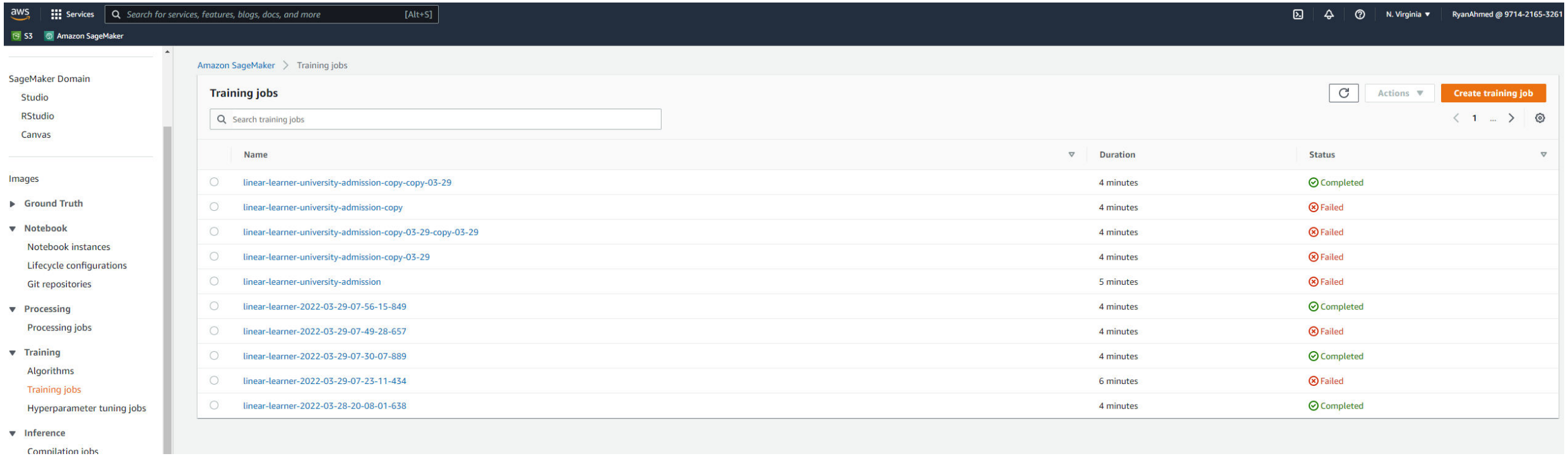
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

CREATE THREE FOLDERS FOR TRAIN, TEST AND OUTPUT. NOTE THAT THE OUTPUT IS THE LOCATION TO STORE THE MODEL ARTIFACTS



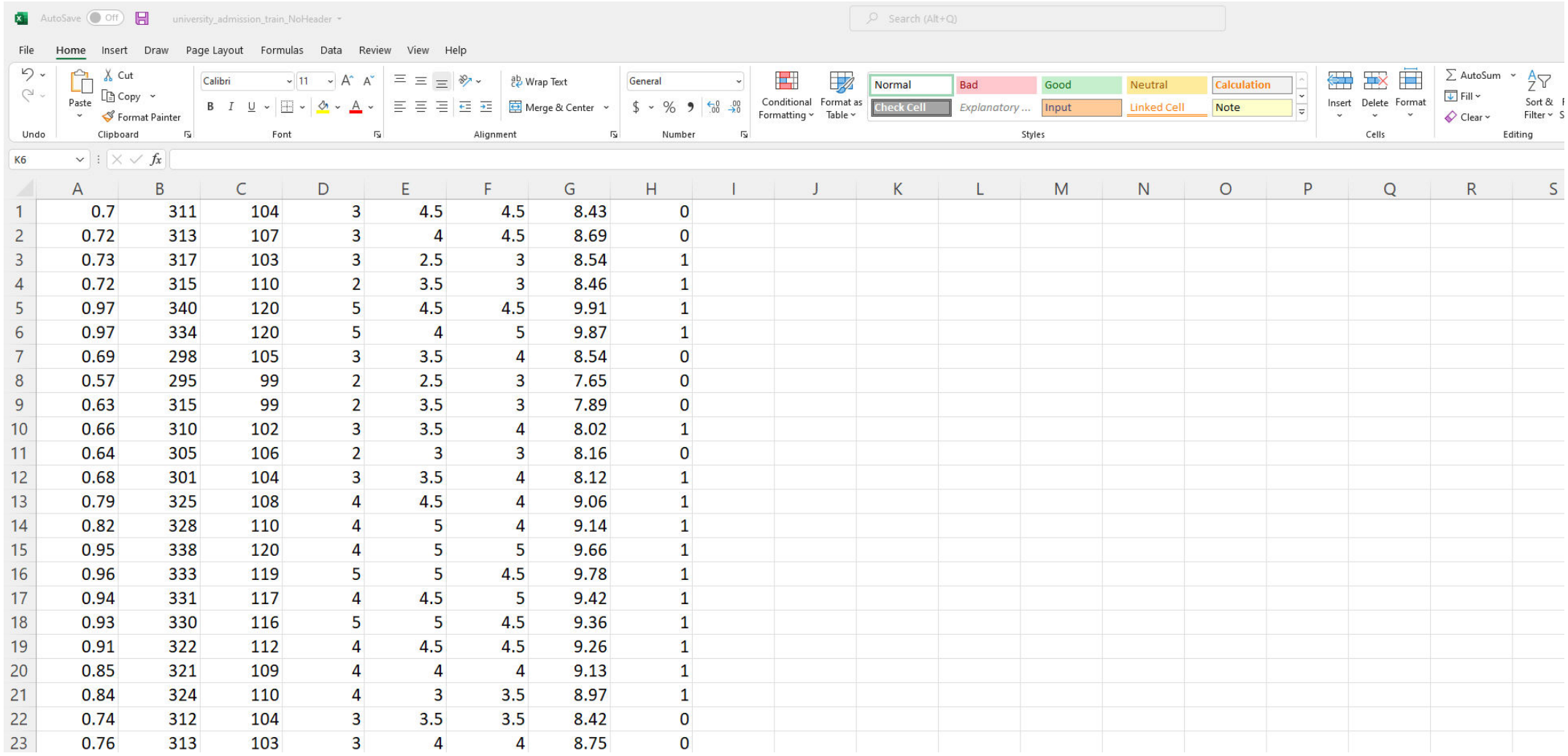
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

NAVIGATE TO SAGEMAKER AND CLICK ON TRAINING JOBS AND THEN CREATE TRAINING JOB



DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

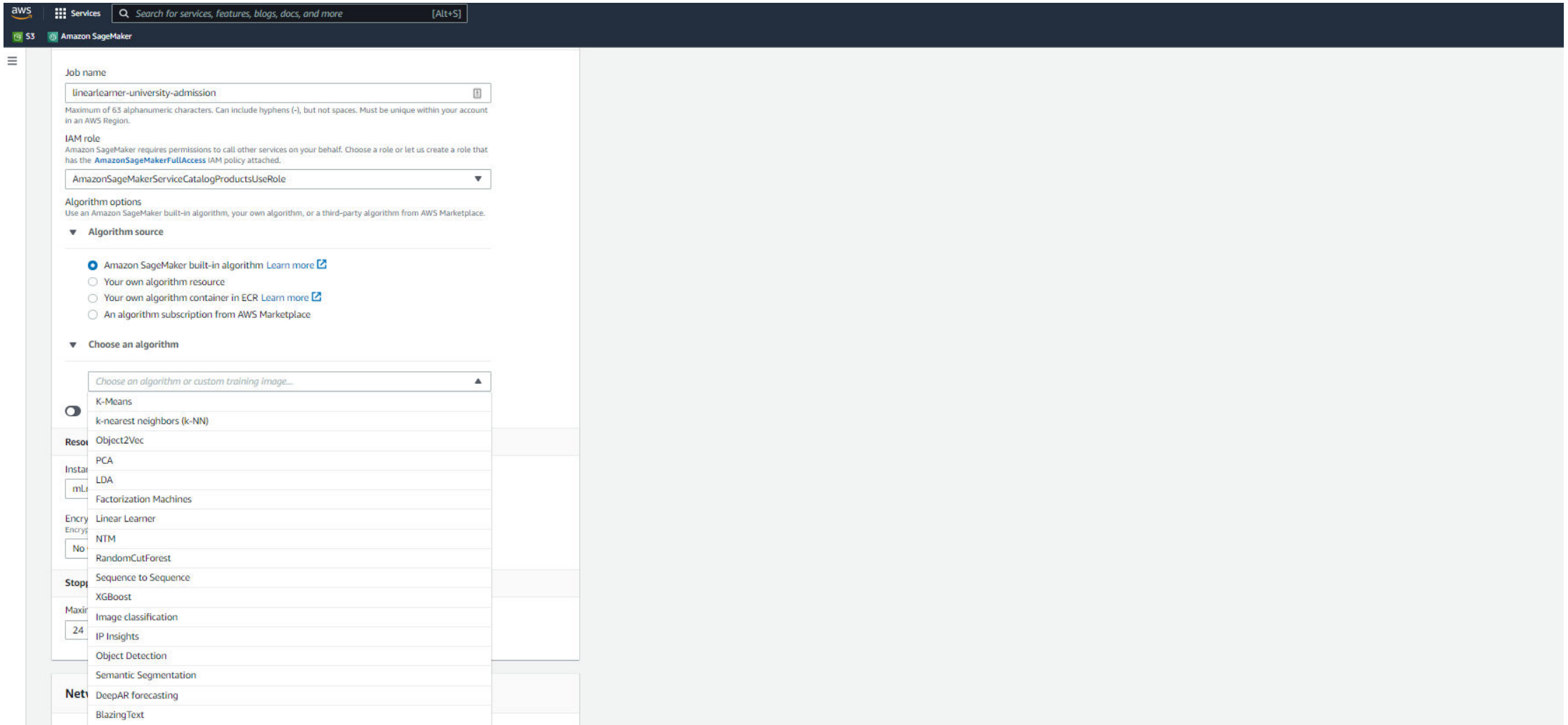
NOTE THAT THE TARGET OUTPUT IS IN THE FIRST COLUMN.
MAKE SURE THAT NO HEADERS ARE PRESENT.



	A	B	C	D	E	F	G	H
1	0.7	311	104	3	4.5	4.5	8.43	0
2	0.72	313	107	3	4	4.5	8.69	0
3	0.73	317	103	3	2.5	3	8.54	1
4	0.72	315	110	2	3.5	3	8.46	1
5	0.97	340	120	5	4.5	4.5	9.91	1
6	0.97	334	120	5	4	5	9.87	1
7	0.69	298	105	3	3.5	4	8.54	0
8	0.57	295	99	2	2.5	3	7.65	0
9	0.63	315	99	2	3.5	3	7.89	0
10	0.66	310	102	3	3.5	4	8.02	1
11	0.64	305	106	2	3	3	8.16	0
12	0.68	301	104	3	3.5	4	8.12	1
13	0.79	325	108	4	4.5	4	9.06	1
14	0.82	328	110	4	5	4	9.14	1
15	0.95	338	120	4	5	5	9.66	1
16	0.96	333	119	5	5	4.5	9.78	1
17	0.94	331	117	4	4.5	5	9.42	1
18	0.93	330	116	5	5	4.5	9.36	1
19	0.91	322	112	4	4.5	4.5	9.26	1
20	0.85	321	109	4	4	4	9.13	1
21	0.84	324	110	4	3	3.5	8.97	1
22	0.74	312	104	3	3.5	3.5	8.42	0
23	0.76	313	103	3	4	4	8.75	0

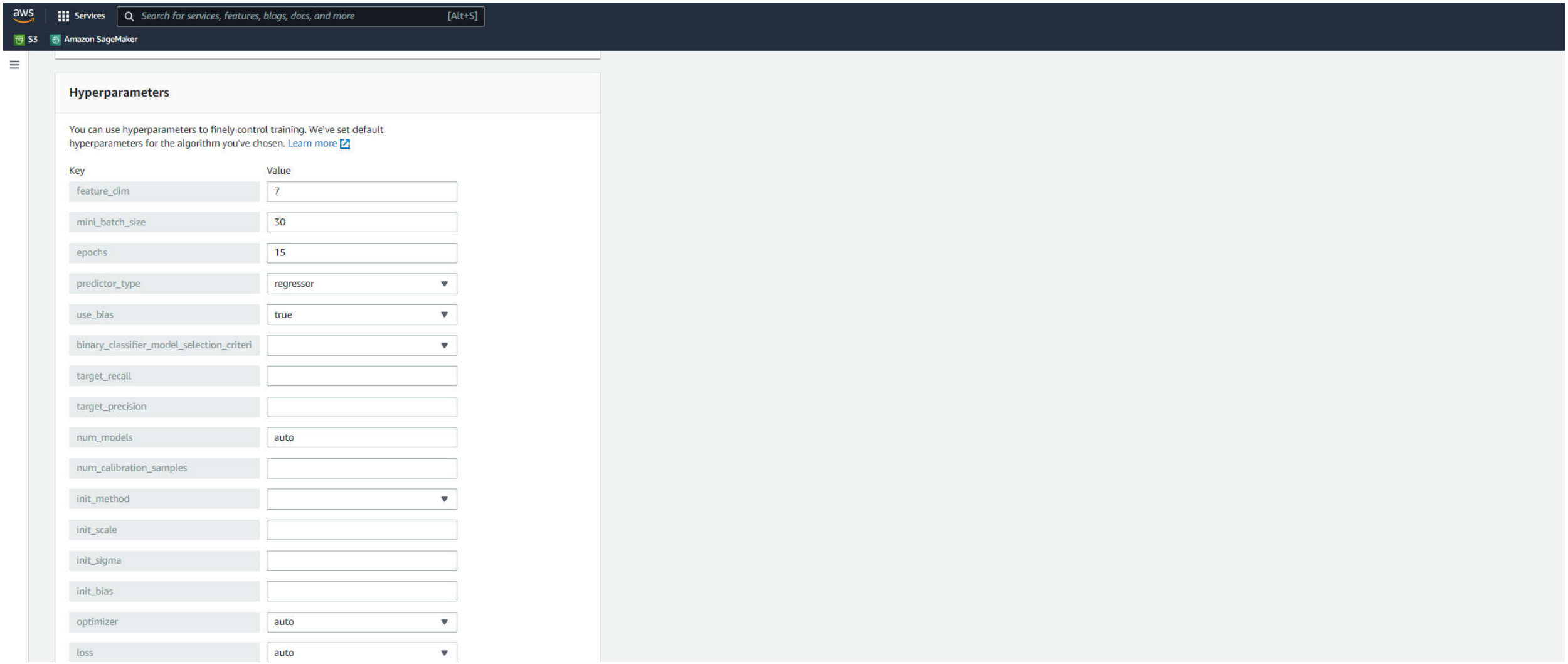
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE A NAME AND CHOOSE A LINEAR LEARNER ALGORITHM



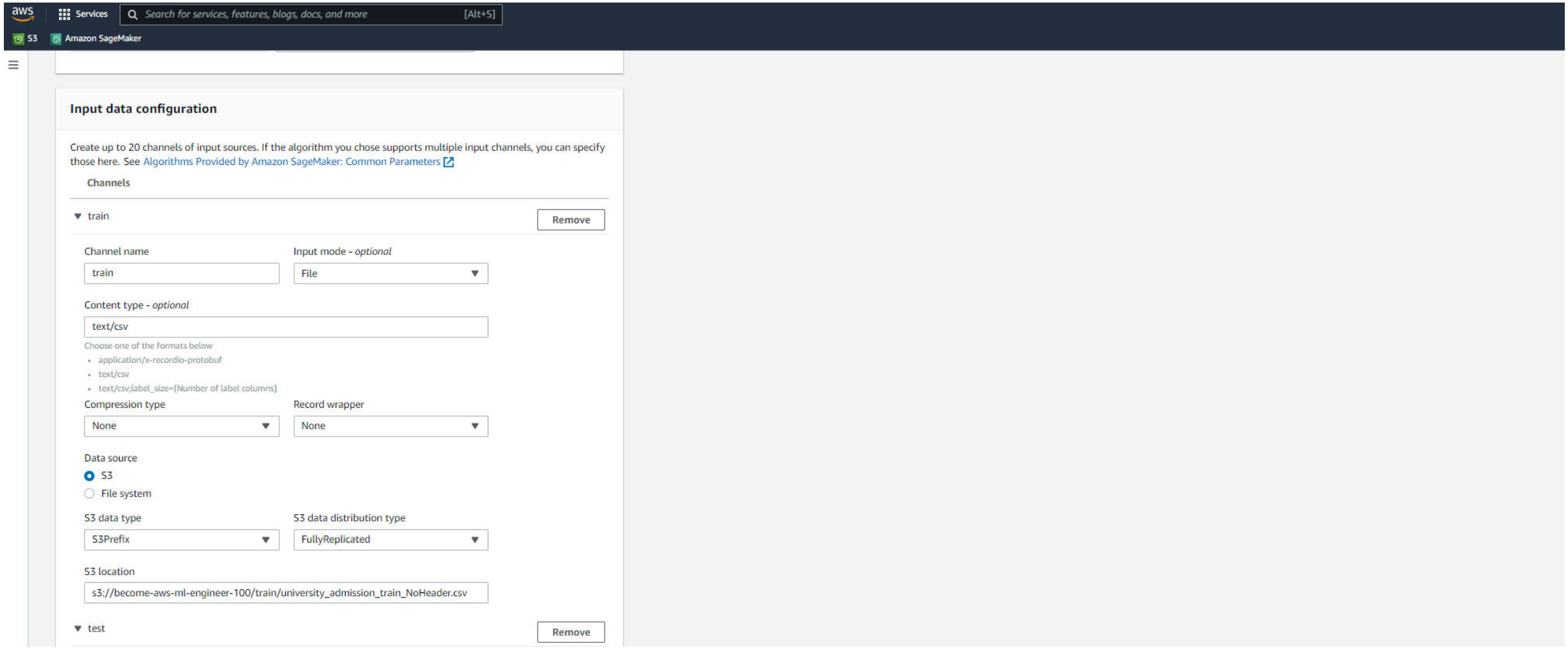
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE FEATURE DIMENSION, MINI_BATCH_SIZE AND PREDICTOR TYPE
(YOU WILL GET AN ERROR IF YOU DON'T PROVIDE THEM.



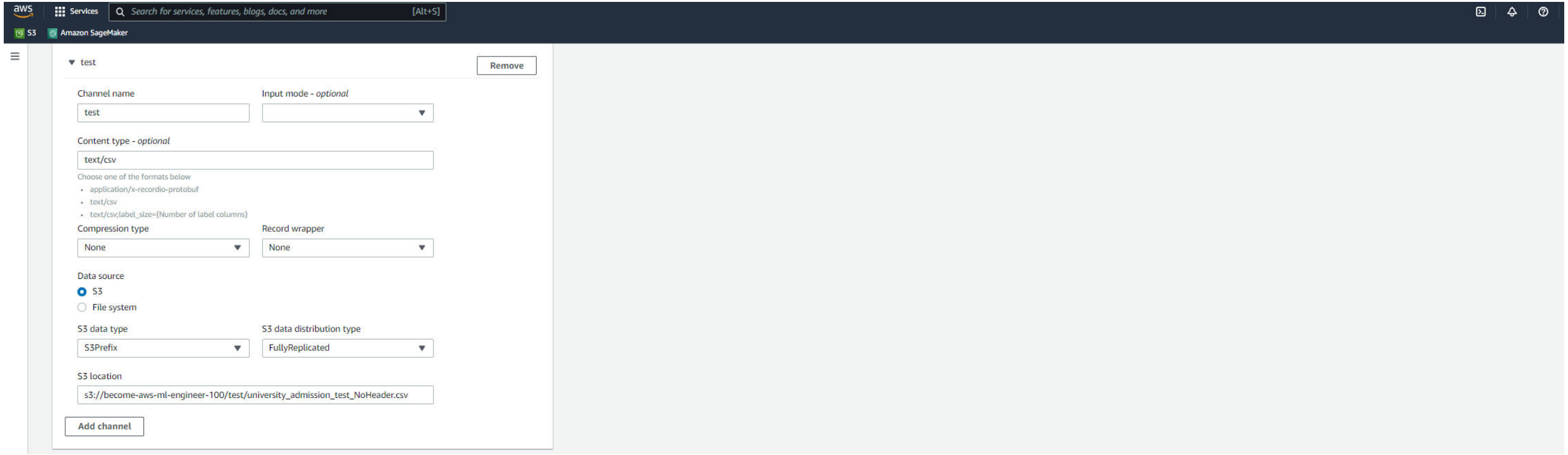
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE THE PATH TO THE TRAINING DATASET IN S3



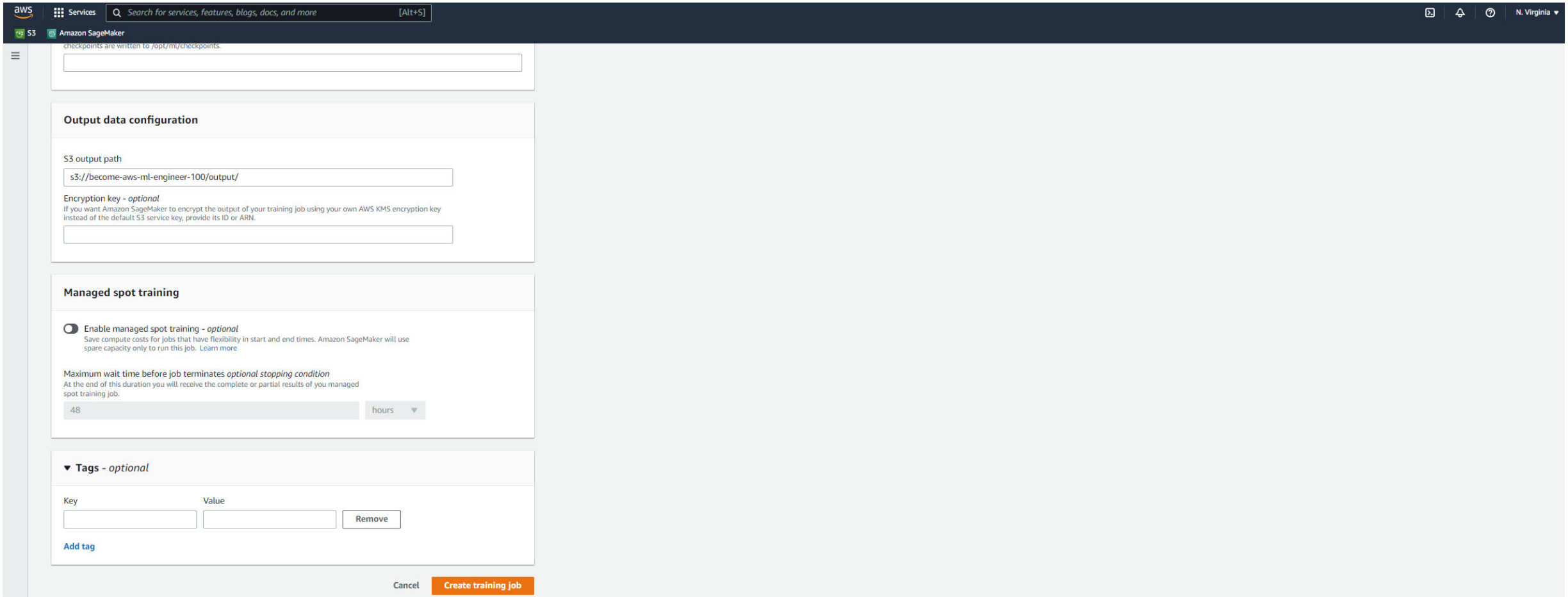
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE THE PATH TO THE TESTING DATASET IN S3



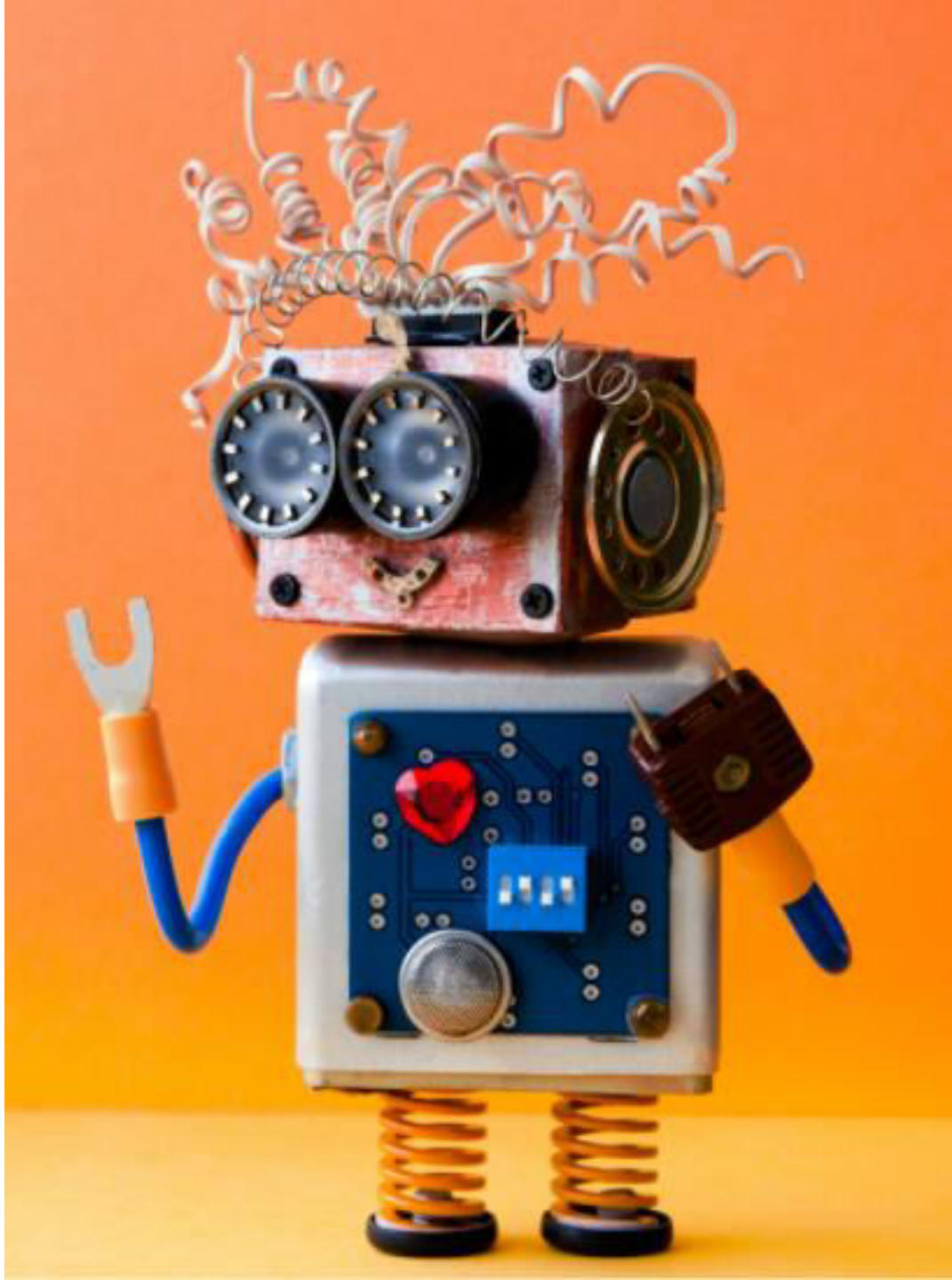
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

SET THE OUTPUT LOCATION AND YOU CAN ENABLE SPOT INSTANCE AS WELL TO SAVE COSTS. CREATE TRAINING JOB.



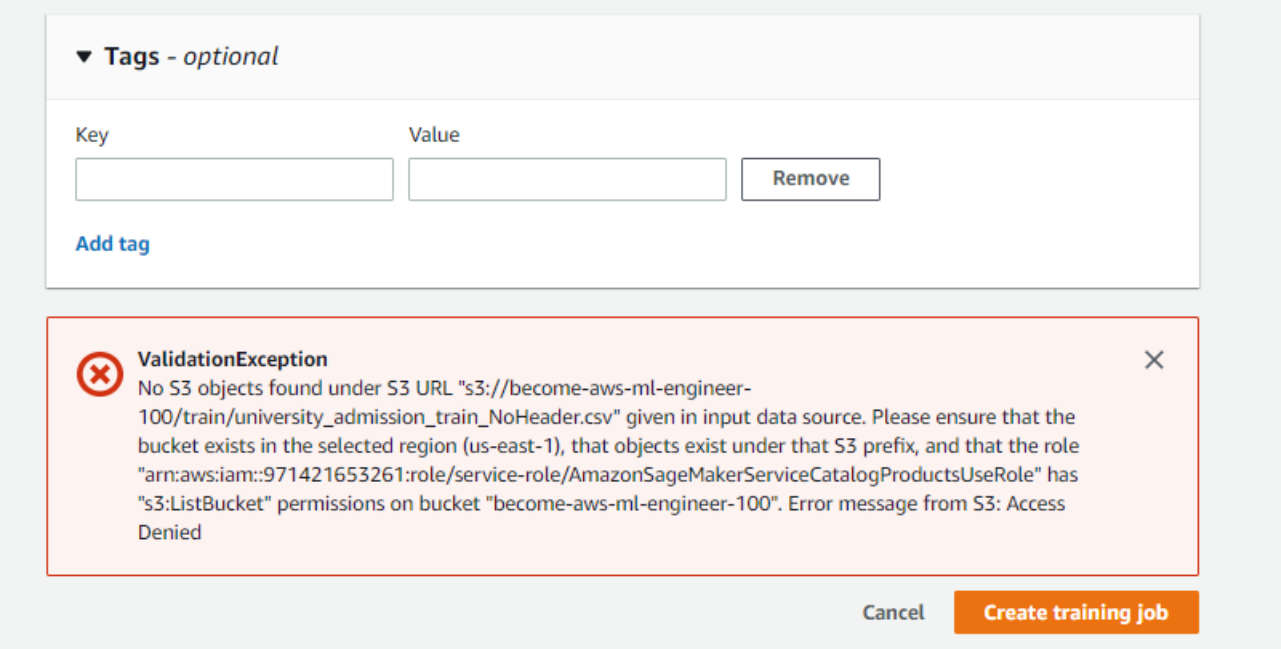
LAUNCH A TRAINING JOB FROM AWS CONSOLE DEMO PART #2

EASY ADVANCED



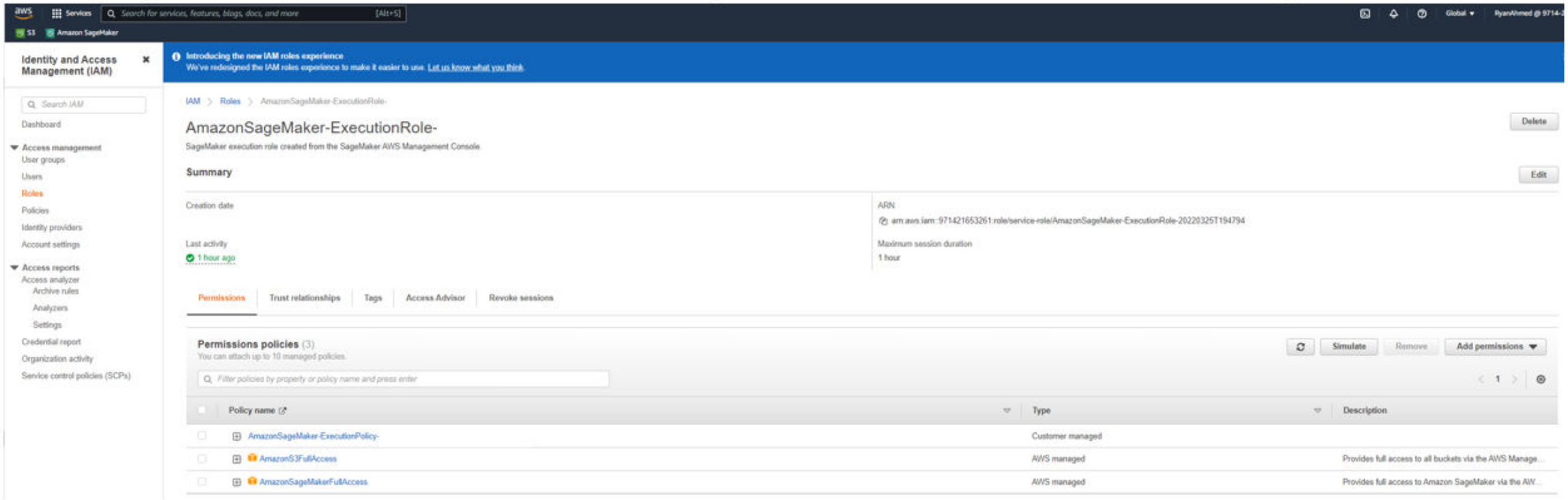
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

NOTE THAT YOU MIGHT GET THIS ERROR MESSAGE. THIS IS BECAUSE SAGEMAKER DOES NOT HAVE ACCESS TO THE S3 BUCKET. LET'S CHANGE THAT!



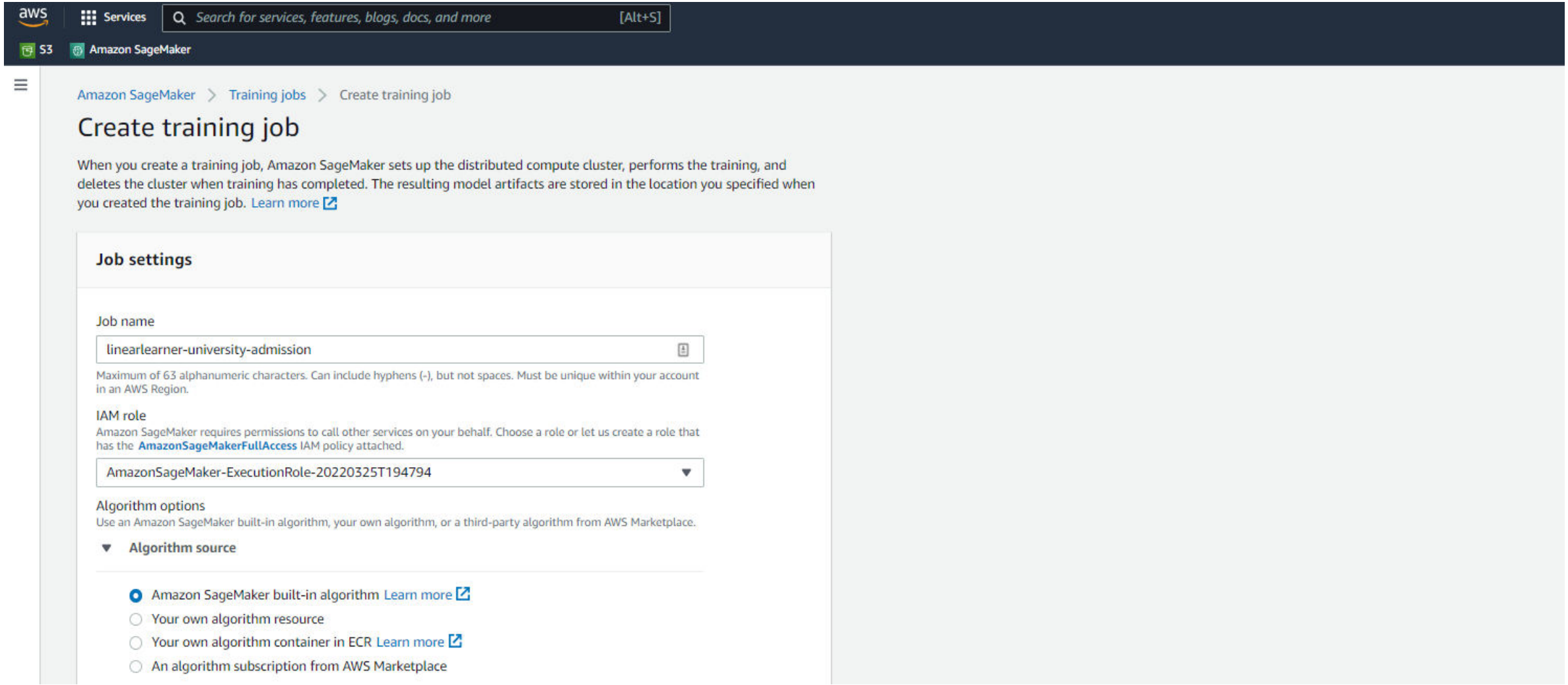
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

NAVIGATE TO IAM AND ATTACH AMAZONS3FULLACCESS TO THE SAGEMAKER EXECUTION ROLE



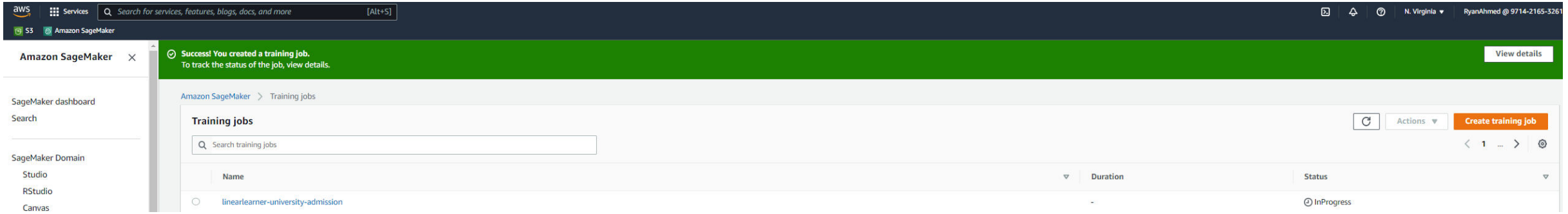
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

SELECT THE SAGEMAKER EXECUTION ROLE THAT HAS FULL S3 ACCESS.
IT SHOULD WORK NOW!



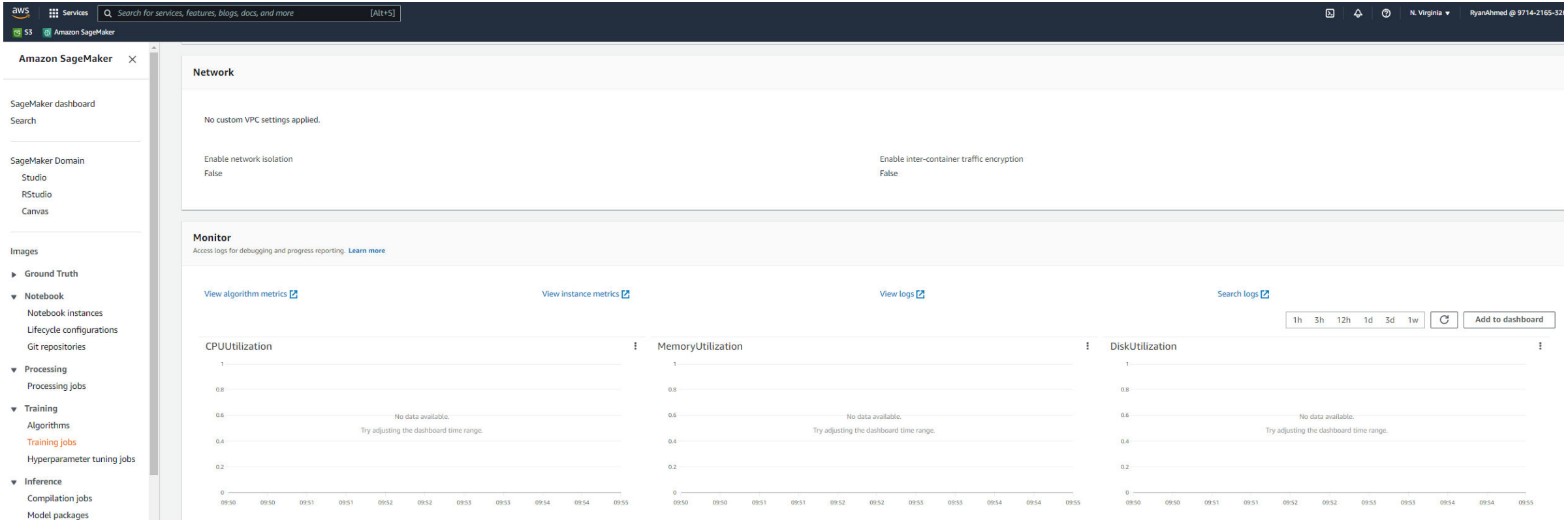
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

TRAINING JOB HAS STARTED!



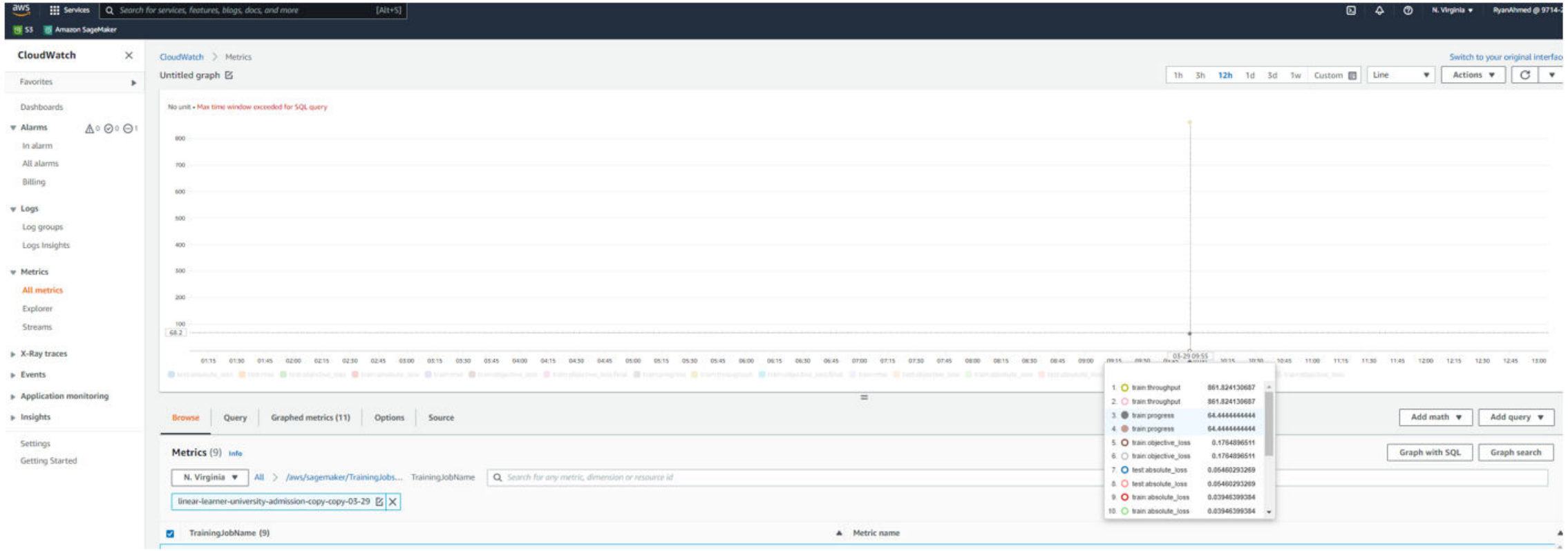
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

ONCE THE TRAINING JOB IS COMPLETE. NAVIGATE TO VIEW ALGORITHM METRICS TO ASSESS TRAINED MODEL PERFORMANCE.



DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

ONCE THE TRAINING JOB IS COMPLETE. NAVIGATE TO VIEW ALGORITHM METRICS TO ASSESS TRAINED MODEL PERFORMANCE. NOTE THAT YOU ARE WATCHING ALL THESE METRICS IN CLOUDWATCH.



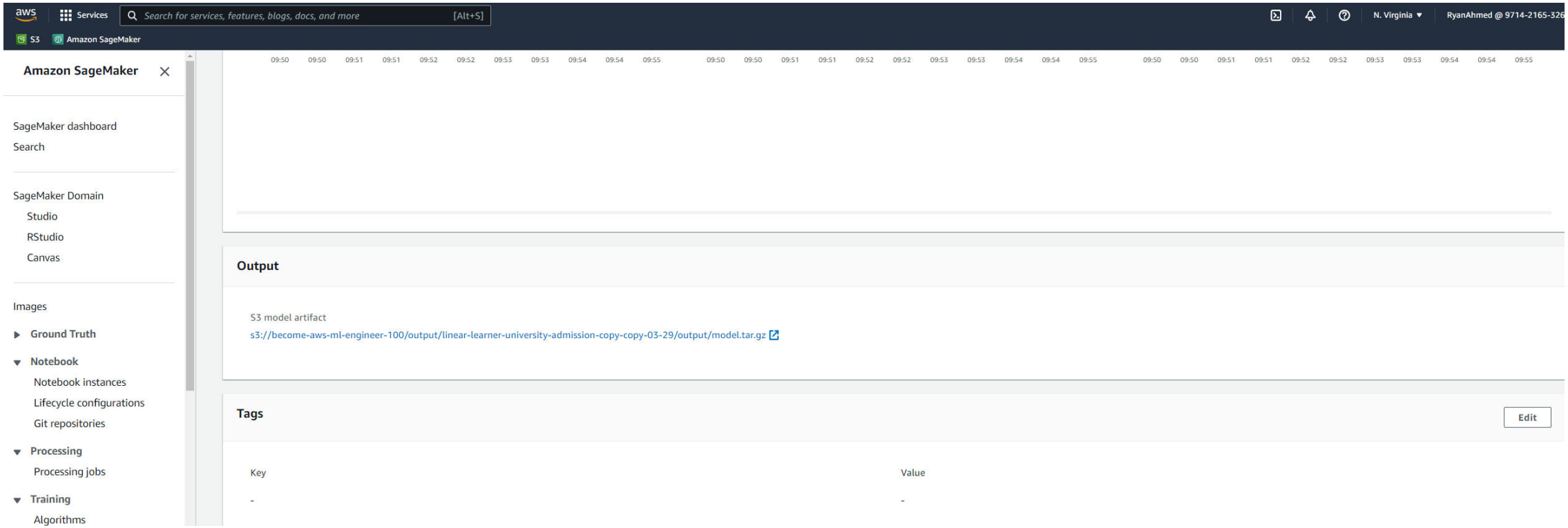
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

CLICK ON VIEW LOGS AND VIEW THE TRAINING JOB METRICS

```
#train_score (algo-1) : ('mse_objective', 0.0030186674197638717)
#train_score (algo-1) : ('mse', 0.0030186674197638717)
#train_score (algo-1) : ('absolute_loss', 0.03946399384008679)
#train_score (algo-1) : ('rmse', 0.05494240092828008)
#train_score (algo-1) : ('r2', 0.8389172107398309)
#train_score (algo-1) : ('mae', 0.0394639939887268)
#quality_metric: host=algo-1, train mse_objective <loss>=0.0030186674197638717
#quality_metric: host=algo-1, train mse <loss>=0.0030186674197638717
#quality_metric: host=algo-1, train absolute_loss <loss>=0.03946399384008679
#quality_metric: host=algo-1, train rmse <loss>=0.05494240092828008
#quality_metric: host=algo-1, train r2 <loss>=0.8389172107398309
#quality_metric: host=algo-1, train mae <loss>=0.0394639939887268
Best model found for hyperparameters: {"optimizer": "adam", "learning_rate": 0.005, "l1": 0.0, "wd": 0.0001, "lr_scheduler_step": 10, "lr_scheduler_factor": 0.99, "lr_scheduler_minimum_lr": 1e-05}
```

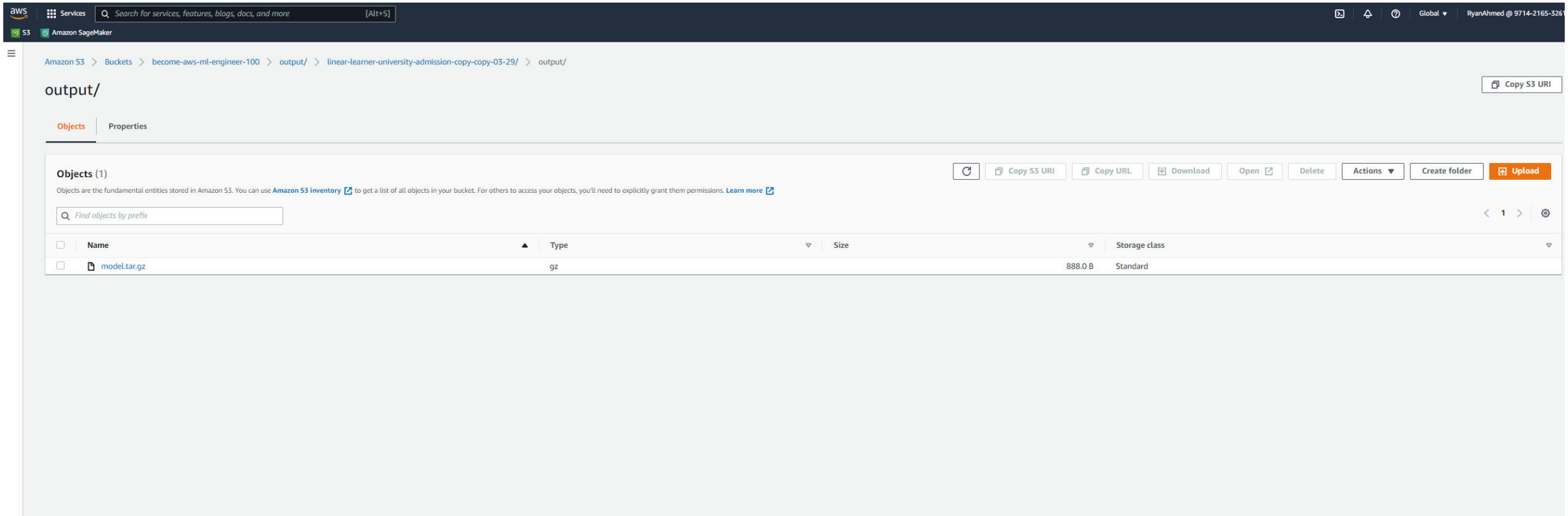
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

ONCE THE TRAINING JOB IS COMPLETE. YOU SHOULD FIND THE TRAINED MODEL ARTIFCATS IN THIS PATH SHOWN BELOW.



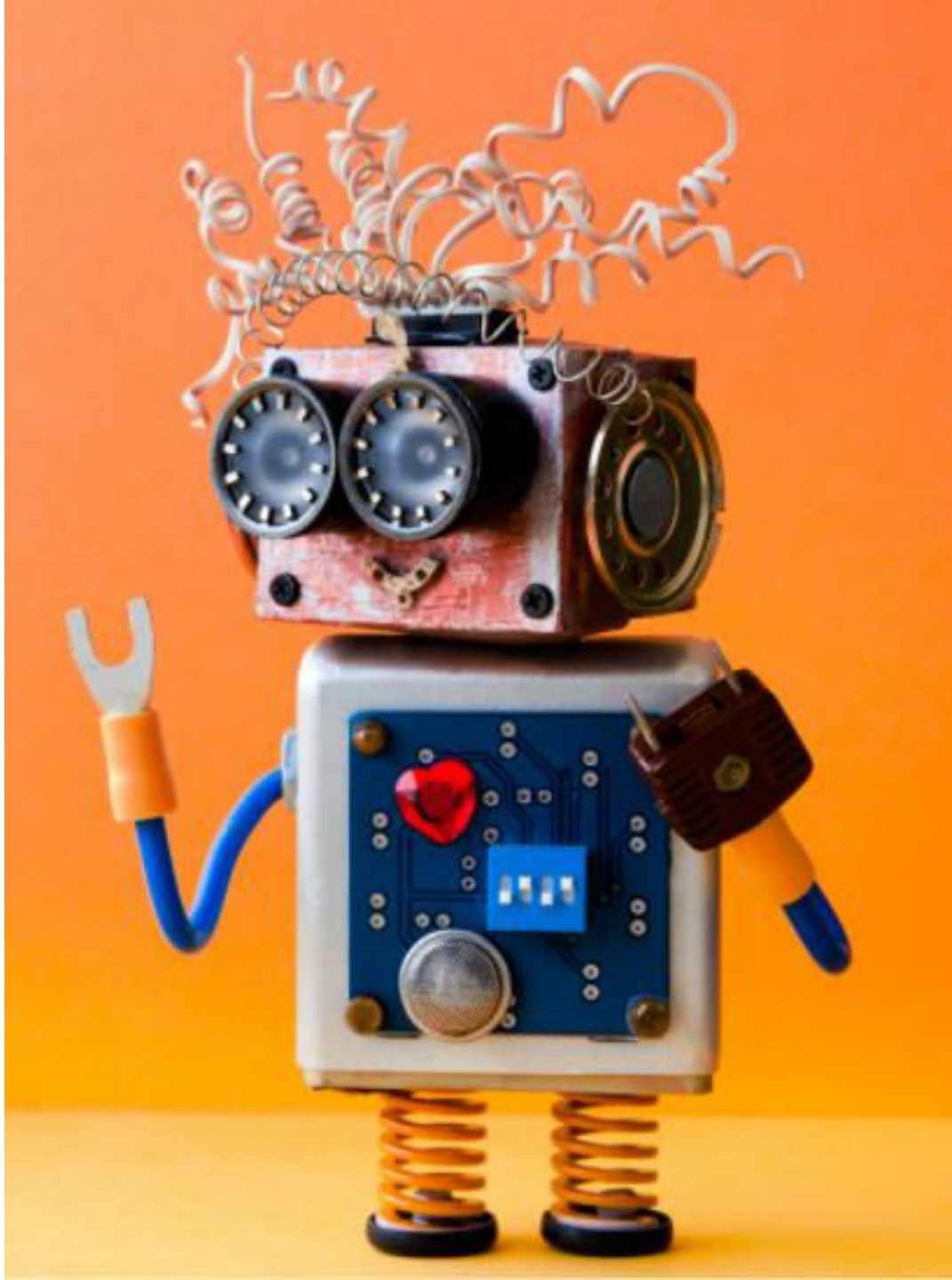
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

TRAINED MODEL OUTPUT IS PLACED IN THE OUTPUT



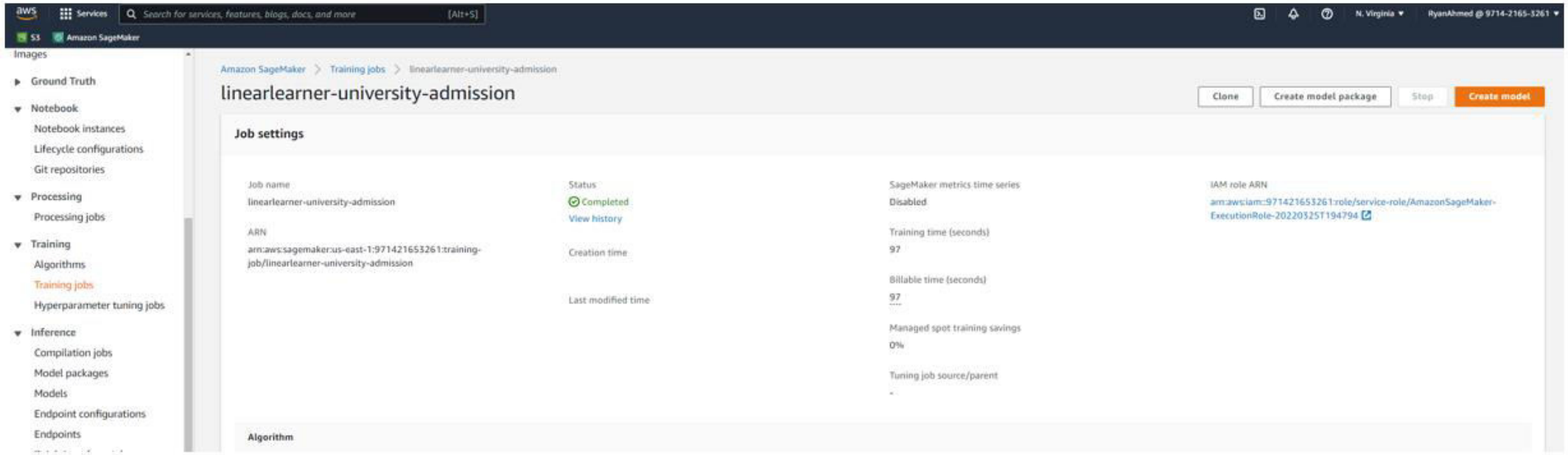
MODEL DEPLOYMENT

EASY ADVANCED



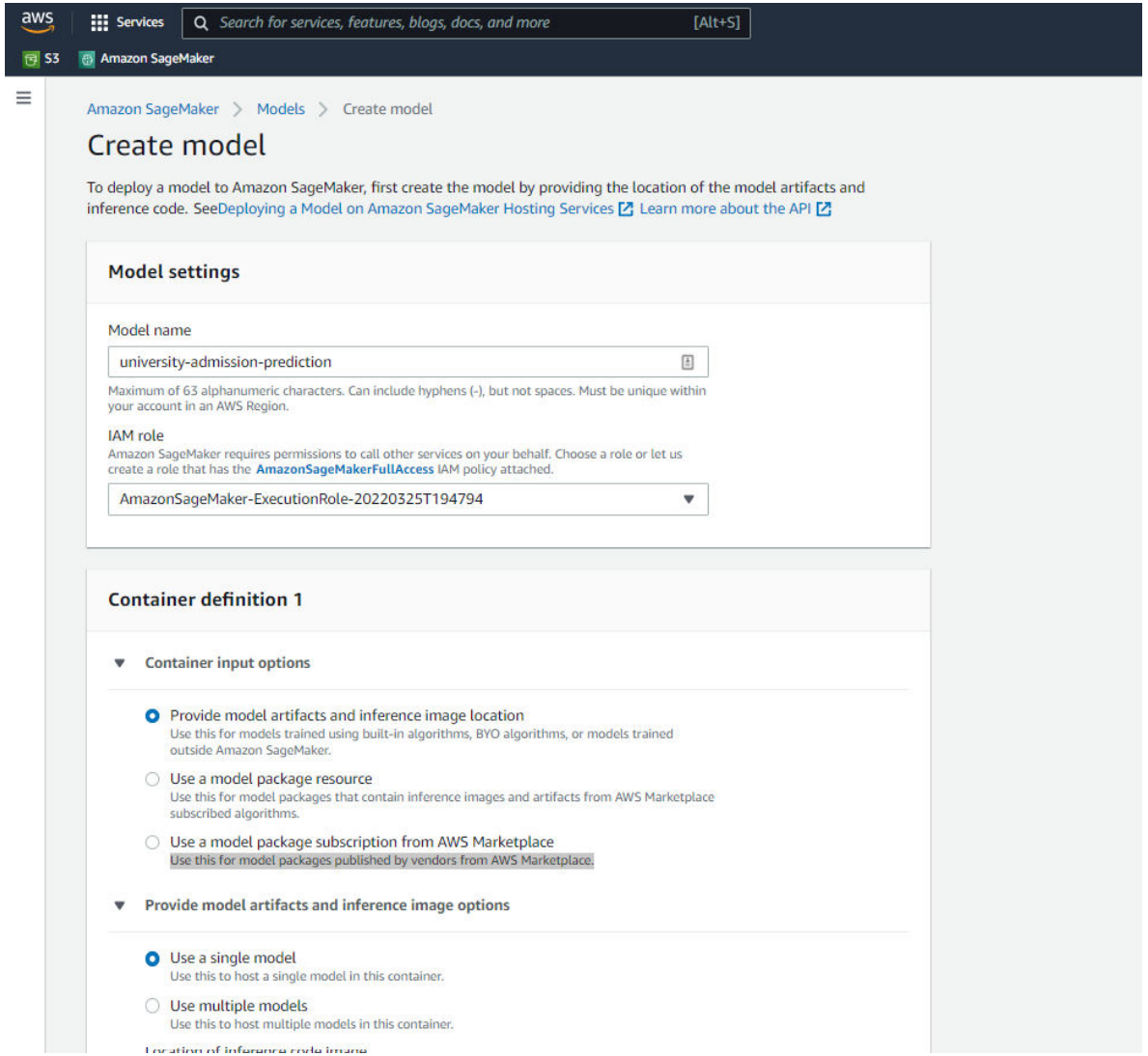
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

ONCE THE TRAINING JOB IS COMPLETE, CLICK ON CREATE MODEL



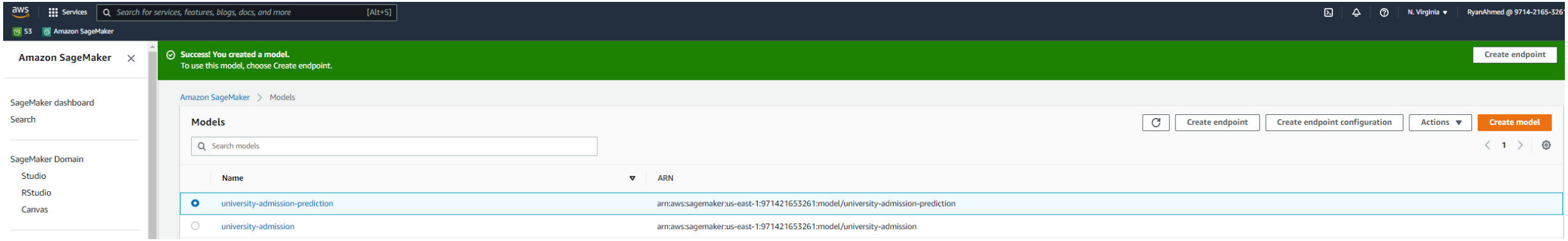
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE A NAME, SELECT THE ROLE AND CLICK CREATE MODEL



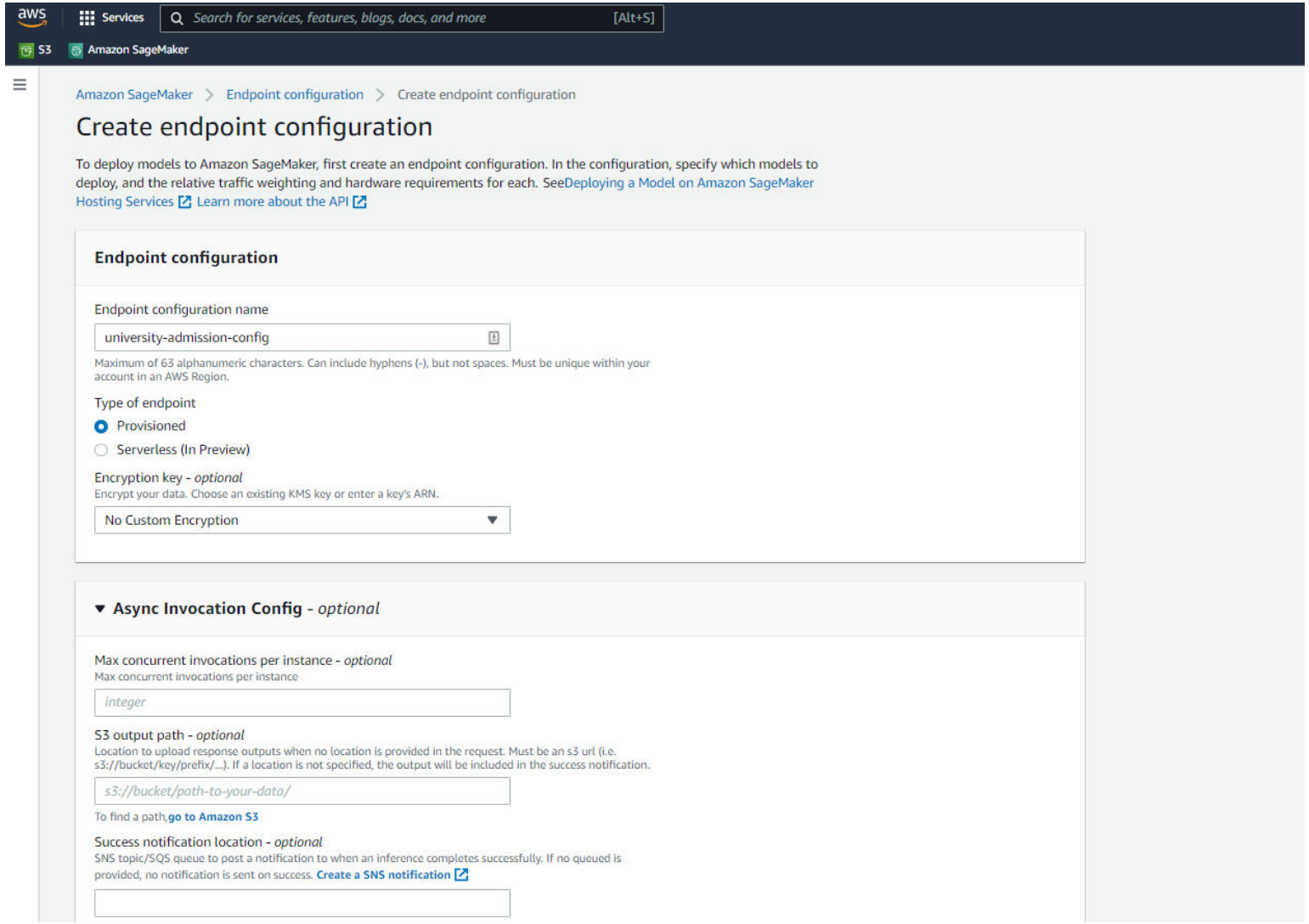
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

CLICK ON CREATE ENDPOINT CONFIGURATION



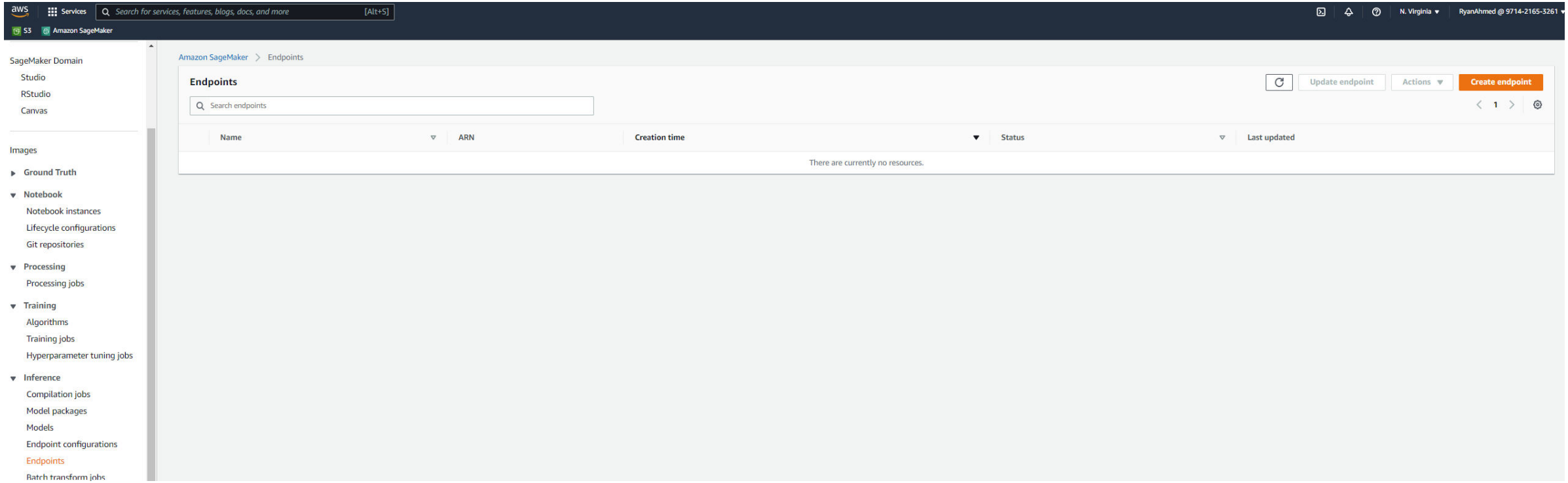
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE A NAME AND CLICK CREATE ENDPOINT CONFIG



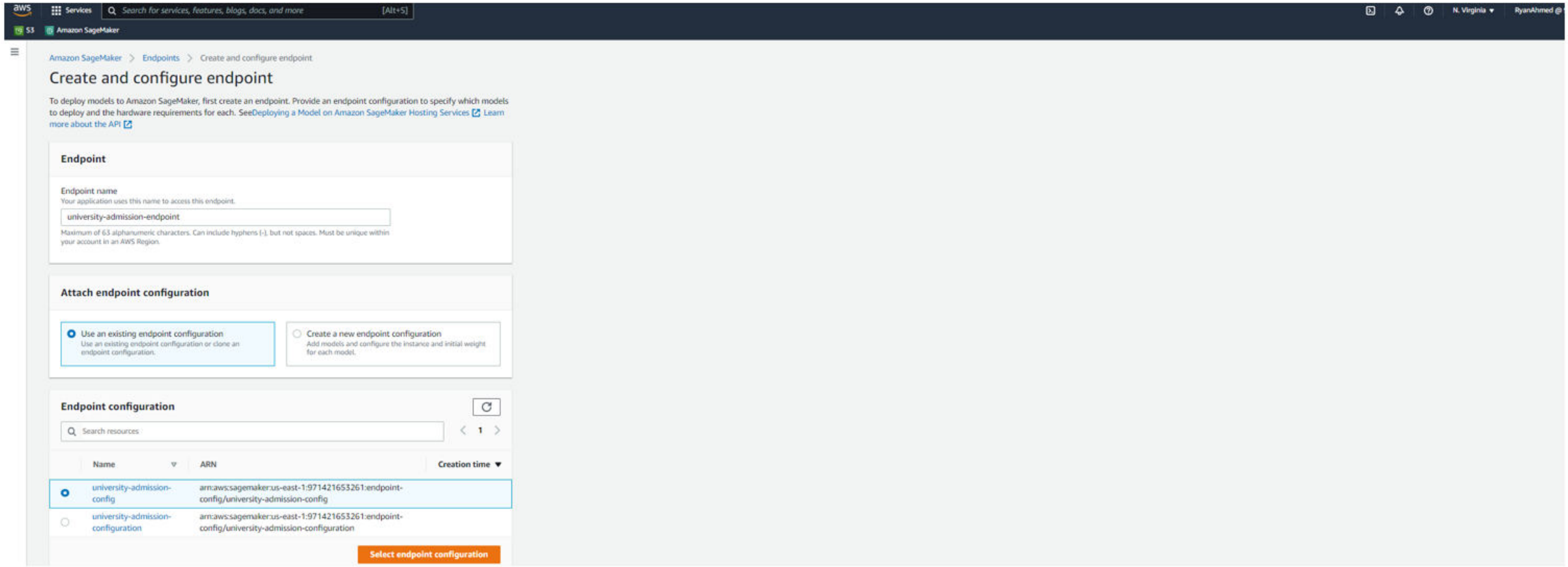
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

NAVIGATE TO ENDPOINTS AND CLICK ON CREATE ENDPOINT



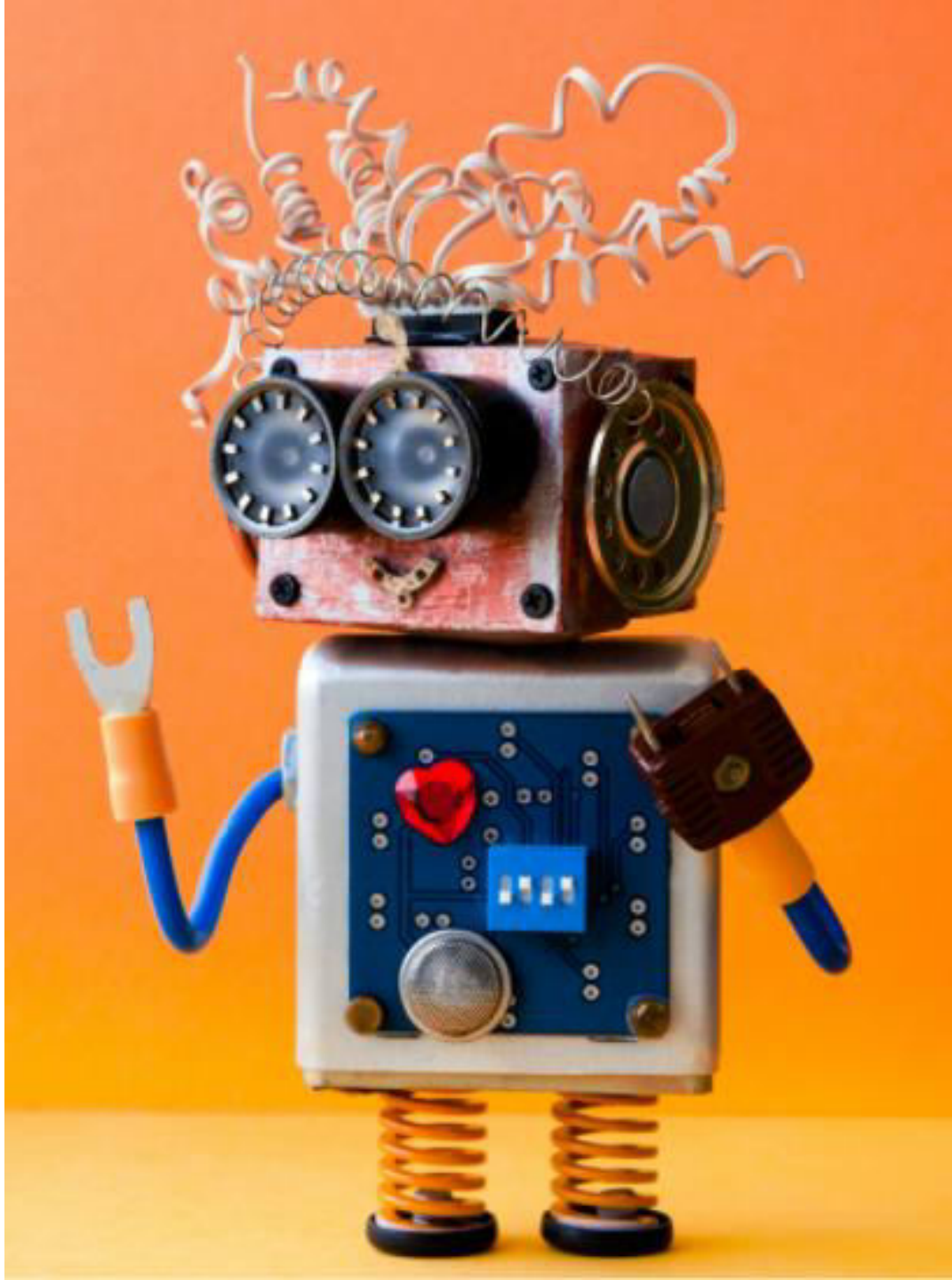
DEMO: LAUNCH TRAINING JOB FROM AWS CONSOLE

PROVIDE AN ENDPOINT NAME AND SELECT THE CONFIGURATION AND CLICK ON CREATE ENDPOINT



FINAL END-OF-DAY CAPSTONE PROJECT

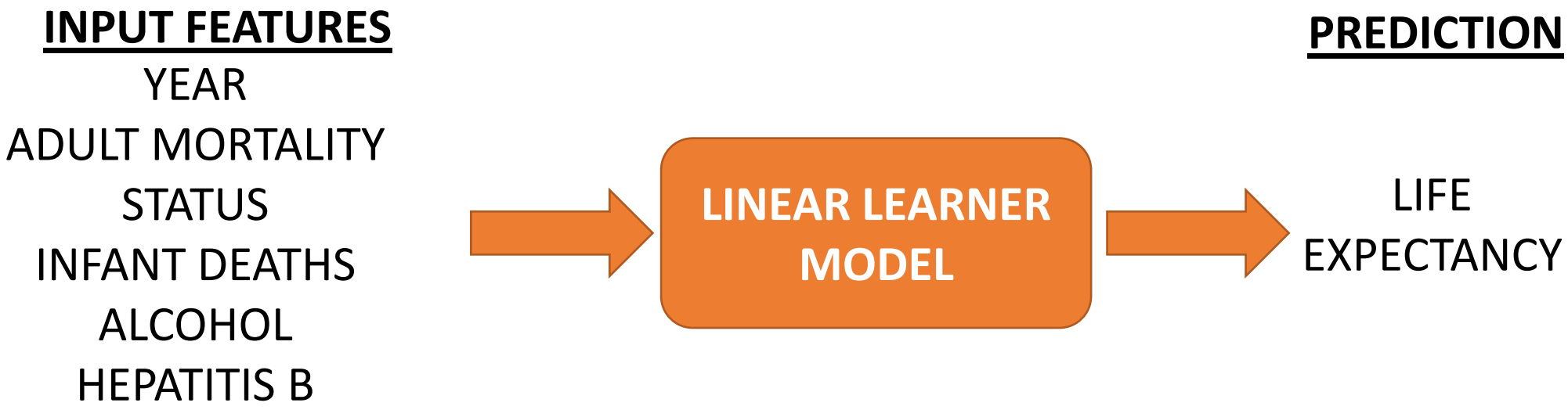
EASY ADVANCED



PROJECT OVERVIEW: LIFE EXPECTANCY PREDICTION

- In this hands-on project, we will train a Linear Regression model to predict life expectancy.
- This data was initially obtained from World Health Organization (WHO) and United Nations Website. Data contains features like year, status, life expectancy, adult mortality, infant deaths, percentage of expenditure, alcohol etc.
- **Tasks:**

1. Upload the dataset Life_Expectancy_test_NoHeader.csv and Life_Expectancy_train_NoHeader.csv to S3
2. Using AWS management console, train a linear leaner model to predict life expectancy.
3. Deploy the model and assess its performance. What's R2?

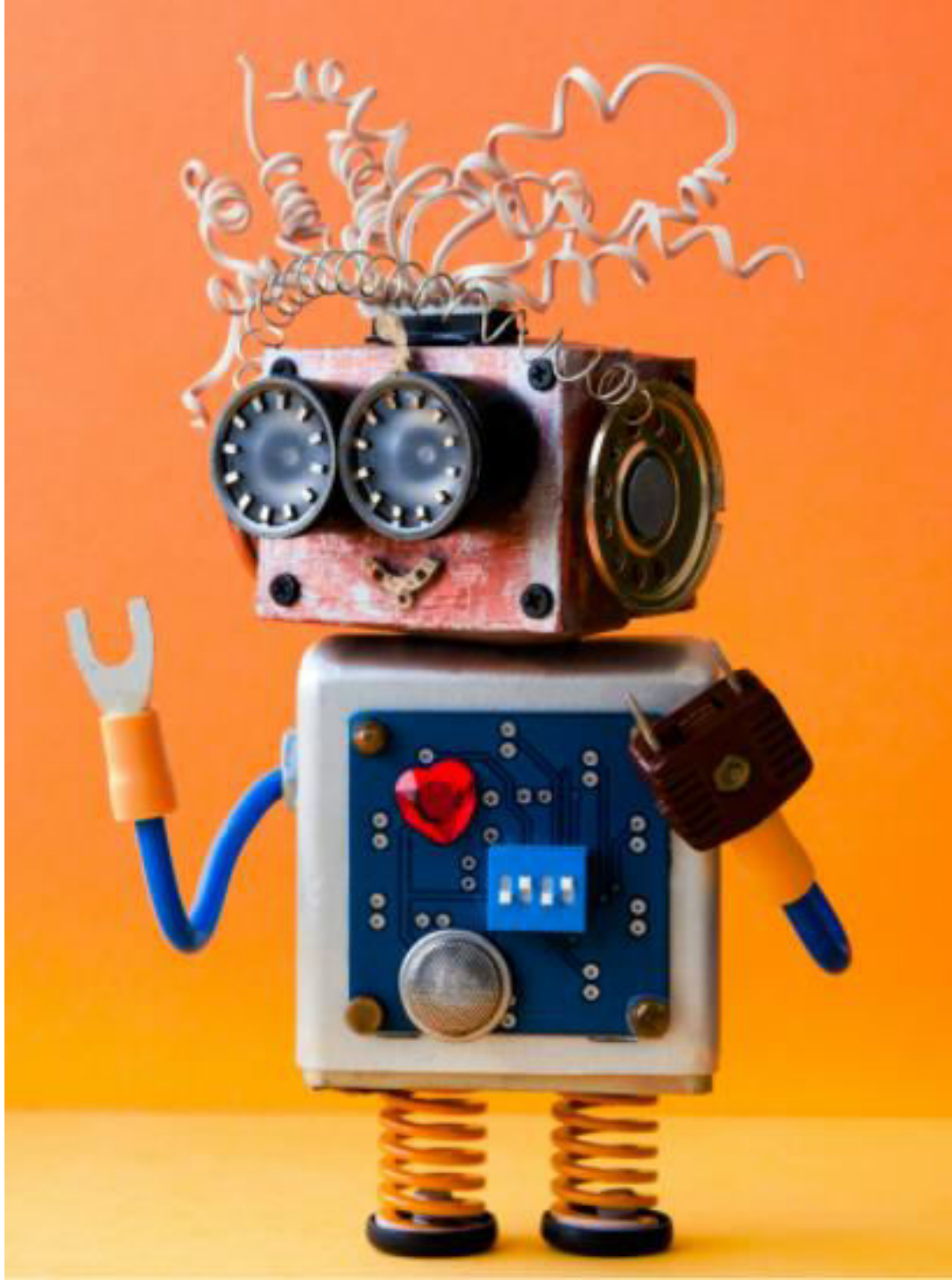


Source: https://www.kaggle.com/jkumarajarshi/life-expectancy-who

FINAL END-OF-DAY CAPSTONE PROJECT SOLUTION

EASY

ADVANCED



PROJECT SOLUTION

```
#test_score (algo-1) : ('mse_objective', 21.298218028175214)
#test_score (algo-1) : ('mse', 21.298218028175214)
#test_score (algo-1) : ('absolute_loss', 3.3841952163110394)
#test_score (algo-1) : ('rmse', 4.614999244655975)
#test_score (algo-1) : ('r2', 0.7977039820516127)
#test_score (algo-1) : ('mae', 3.3841952099425963)
#quality_metric: host=algo-1, test mse_objective <loss>=21.298218028175214
#quality_metric: host=algo-1, test mse <loss>=21.298218028175214
#quality_metric: host=algo-1, test absolute_loss <loss>=3.3841952163110394
#quality_metric: host=algo-1, test rmse <loss>=4.614999244655975
#quality_metric: host=algo-1, test r2 <loss>=0.7977039820516127
#quality_metric: host=algo-1, test mae <loss>=3.3841952099425963
"EndTime": 1651035781.2368584, "Dimensions": {"Algorithm": "Linear Learner", "Host": "algo-1", "Operation": "training"}, "Metrics": {"initialize.time": {"sum": 848.08659…
```