Predicting Success Rates of Space X Rockets

James J. Heffers

University of Michigan - Ann Arbor

heffers@umich.edu

12/14/2022

Introduction

- SpaceX is a company that launches many various payloads into orbit. They have rockets that can successfully land after the initial launch, allowing them to be reused, saving millions in costs of having to rebuilt new rockets each time.
- We want to make a model that can help us determine if a new rocket will be able to land successfully based on the given data.
- This presentation just shares visualizations, results, and insights. All processes/code can be found in the notebooks in the accompanying GitHub repository! Link provided at end for convenience.

Methodology

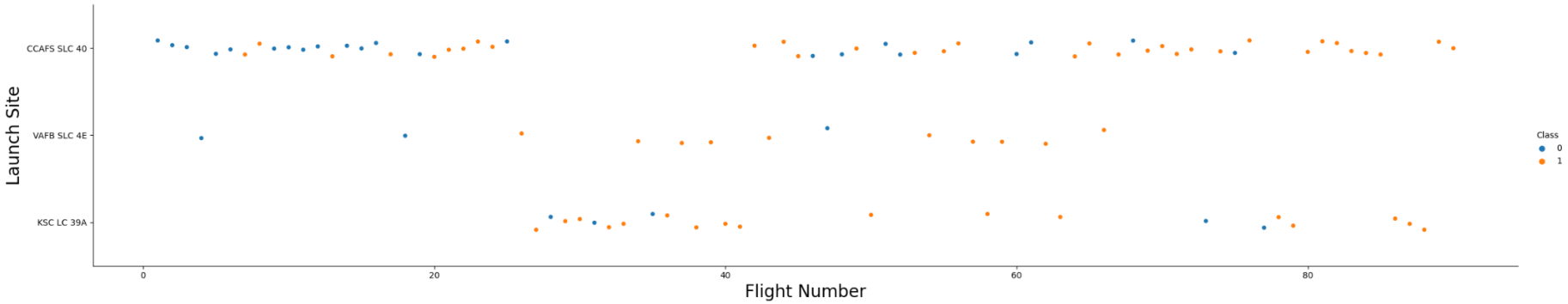
- Data is collected using SpaceX API and html scraping the SpaceX Wiki page.
- Use SQL magic in python to organize and explore the data.
- Perform exploratory data analysis using visualizations.
- Perform predictive analysis using classification models and tuning hyperparameters.
- Deploy the model as a real time inference pipeline on Microsoft Azure Platform.

EDA Insights

- In the upcoming few slides we will see insights gained from the EDA. First we start with various visualizations to help us see the relationships between the different features.
- We see that the payload mass, the orbit type, and the year of the launch all impact the success rates of a landing.

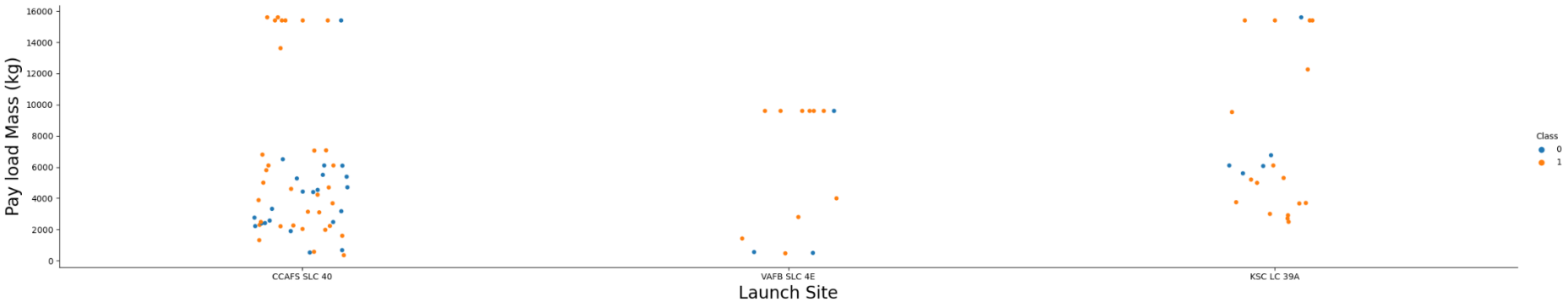
Flight Number vs Launch Site

```
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



Payload vs Launch Site

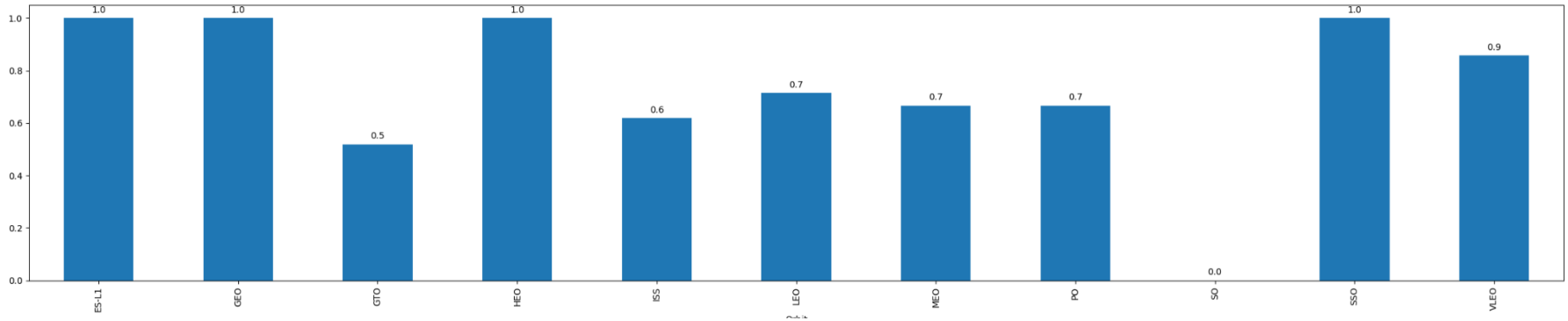
```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
sns.catplot(y="PayloadMass", x="LaunchSite", hue="Class", data=df, aspect = 5)
plt.xlabel("Launch Site",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



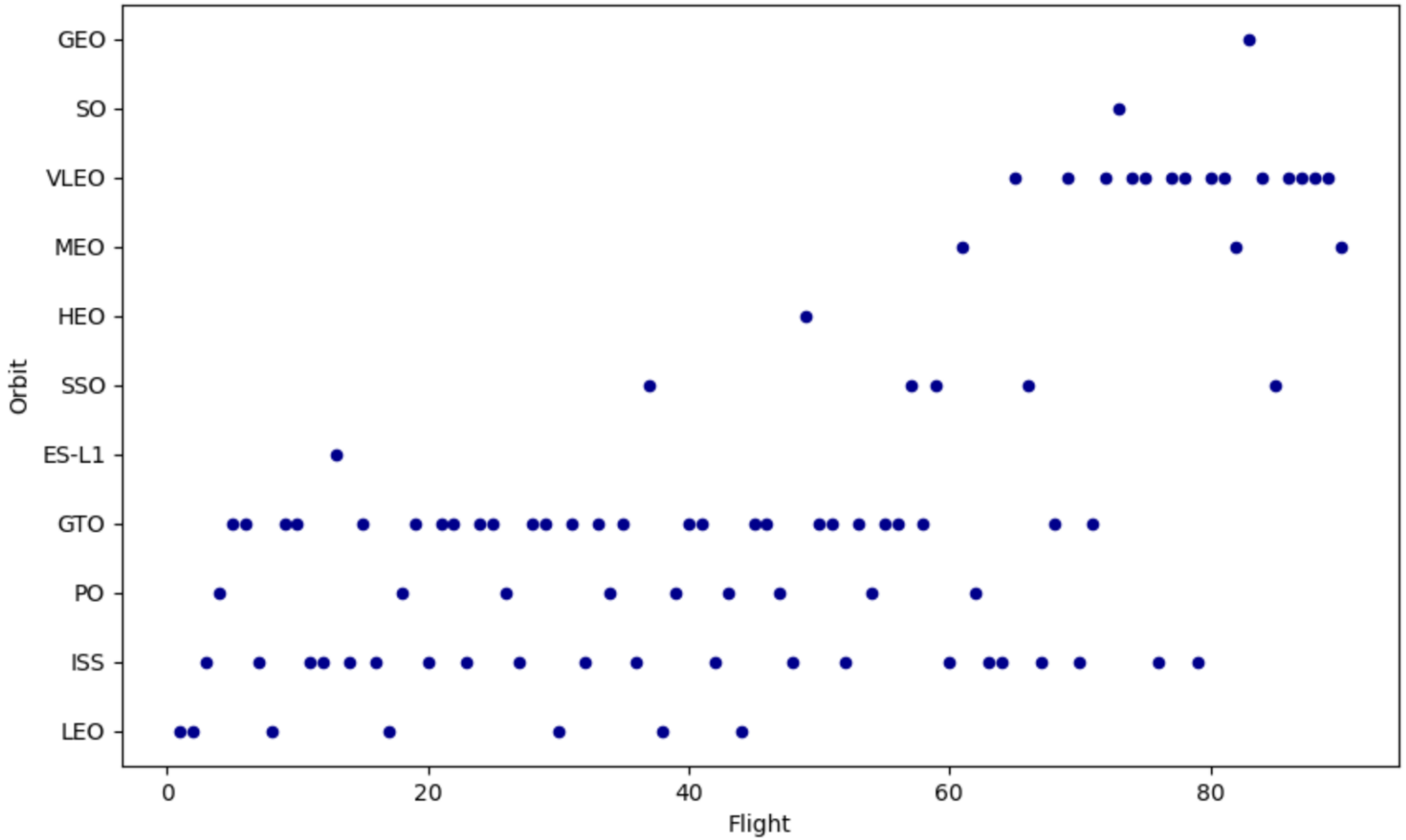
Success Rate vs Orbit Type

```
splot=df.groupby(['Orbit']).mean()['Class'].plot(kind='bar')
for p in splot.patches:
    splot.annotate(format(p.get_height(), '.1f'),
                   (p.get_x() + p.get_width() / 2., p.get_height()),
                   ha = 'center', va = 'center',
                   xytext = (0, 9),
                   textcoords = 'offset points')


plt.show()
```

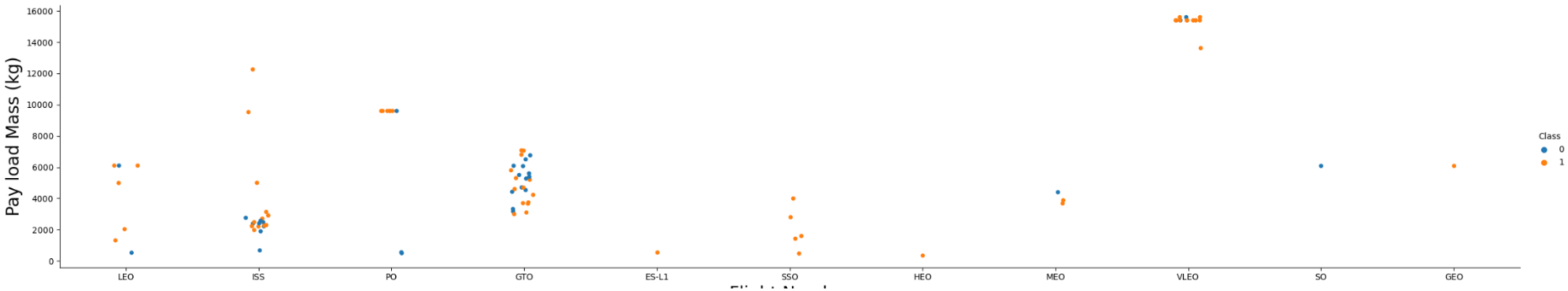


Flight Number vs Orbit Type



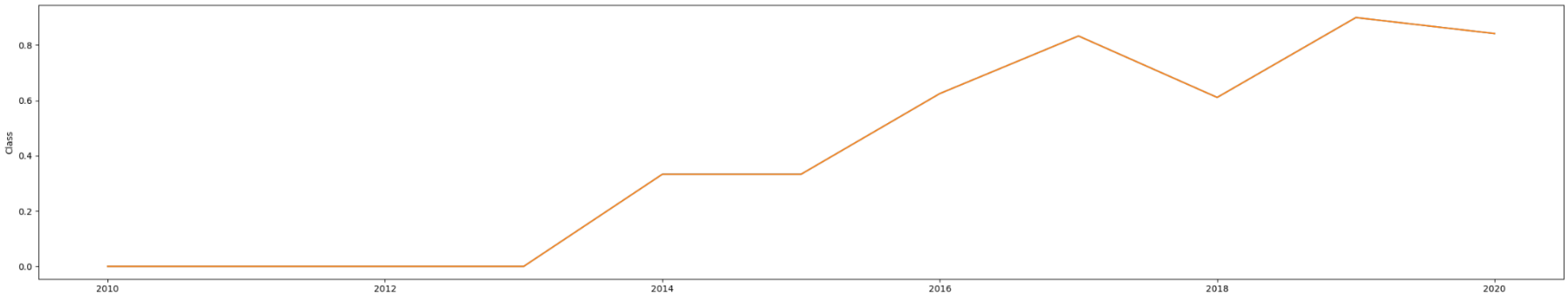
Payload vs Orbit Type

```
sns.catplot(y="PayloadMass", x="Orbit", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



Success Rate Yearly Trend

```
sns.lineplot(x = df['Year'].unique() , y = df.groupby(['Year'])['Class'].mean())
plt.show()
```



Average Payload Mass

- Now we will use SQL magic to do some numerical analysis! We start with finding the average payload mass over all the launches.

```
%sql SELECT Booster_Version, AVG(PAYLOAD_MASS__KG_) as Ave_Payload_mass FROM SPACEXTBL where Booster_Version LIKE 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Ave_Payload_mass
F9 v1.1 B1003	2534.6666666666665

First Successful Landing

```
%sql SELECT min("DATE") from SPACEXTBL where "LANDING _OUTCOME" ='Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

min("DATE")
01-05-2017

Successful Landings With Payload Between 4000-6000kg

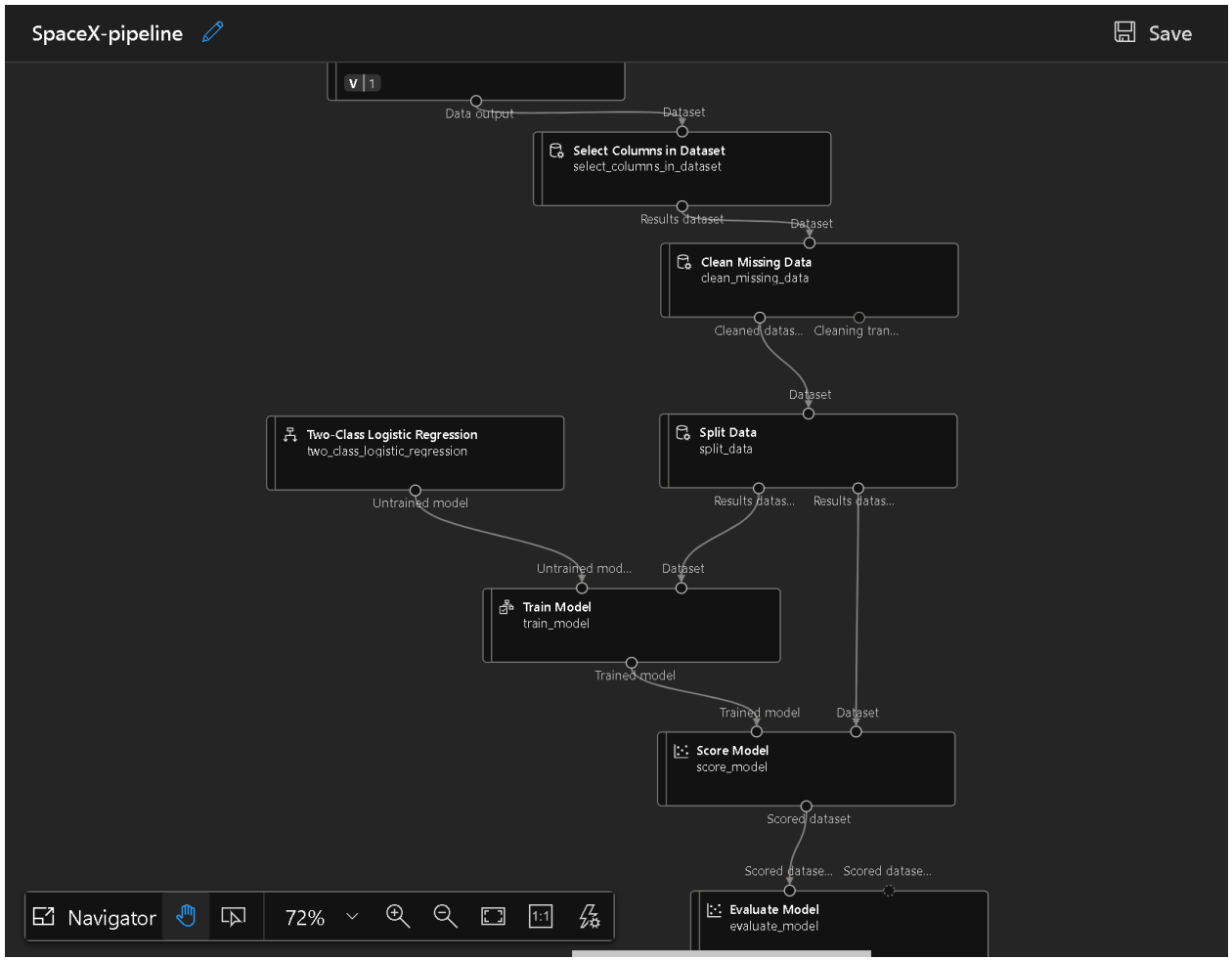
```
%sql SELECT Booster_Version,Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE "MISSION_OUTCOME"="Success" AND "PAYLOAD_MASS__KG_" < 6000 AND "PAYLOAD_MASS_
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 v1.1	AsiaSat 8	4535
F9 v1.1 B1011	AsiaSat 6	4428
F9 v1.1 B1014	ABS-3A Eutelsat 115 West B	4159
F9 v1.1 B1016	Turkmen 52 / MonacoSAT	4707
F9 FT B1020	SES-9	5271
F9 FT B1022	JCSAT-14	4696
F9 FT B1026	JCSAT-16	4600
F9 FT B1030	EchoStar 23	5600
F9 FT B1021.2	SES-10	5300
F9 FT B1032.1	NROL-76	5300
F9 B4 B1040.1	Boeing X-37B OTV-5	4990
F9 FT B1031.2	SES-11 / EchoStar 105	5200
F9 FT B1032.2	GovSat-1 / SES-16	4230
F9 B4 B1040.2	SES-12	5384
F9 B5 B1046.2	Merah Putih	5800
F9 B5 B1047.2	Es hail 2	5300
F9 B5 B1048.3	Nusantara Satu, Beresheet Moon lander, S5	4850
F9 B5 B1051.2	RADARSAT Constellation, SpaceX CRS-18	4200
F9 B5B1060.1	GPS III-03, ANASIS-II	4311
F9 B5 B1058.2	ANASIS-II, Starlink 9 v1.0	5500
F9 B5B1062.1	GPS III-04 , Crew-1	4311

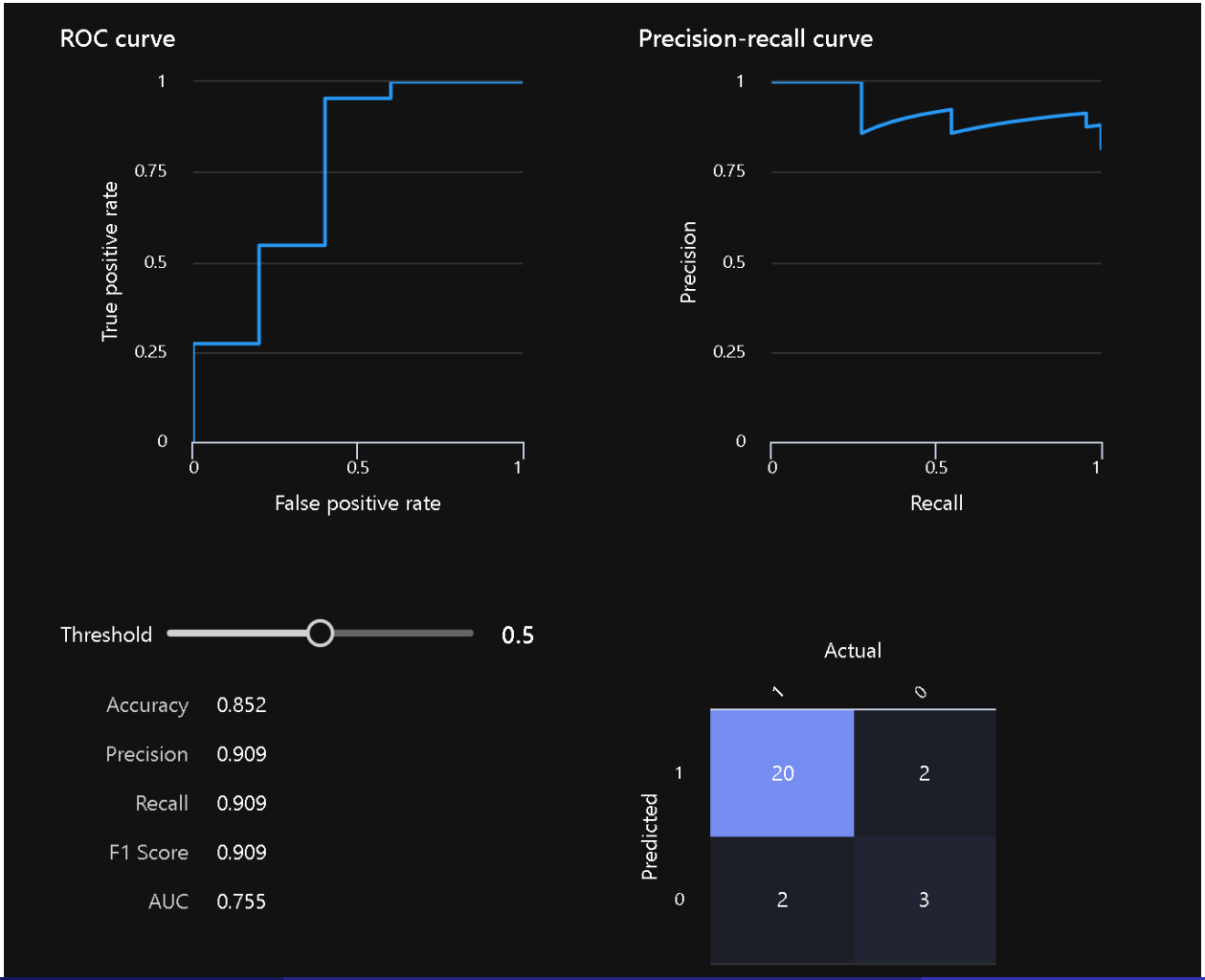
Modeling

- Now we will construct some models. We start by using Azure Machine Learning Studio to create a pipeline for some quick no-code insights.



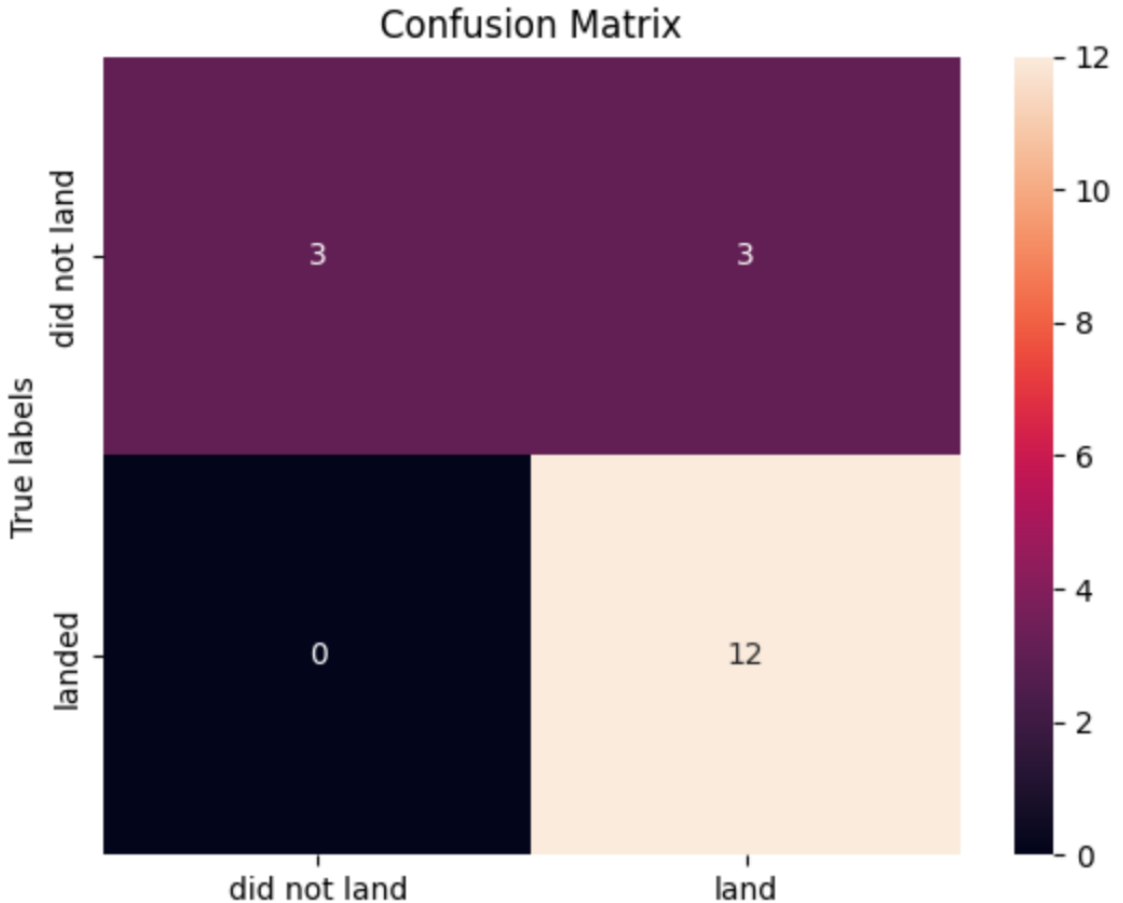
Modeling

- Checking the job output, it seems a classification model would be promising! However we will fine-tune this with some coding. Below is a sample of some of the outputs from the pipeline we constructed.



Modeling

- We run several other models in addition to logistic regression, such as a decision tree, SVM, and K nearest neighbors. We see that each model has a very similar accuracy score of roughly 0.833. Below is the confusion matrix corresponding to the SVM model.

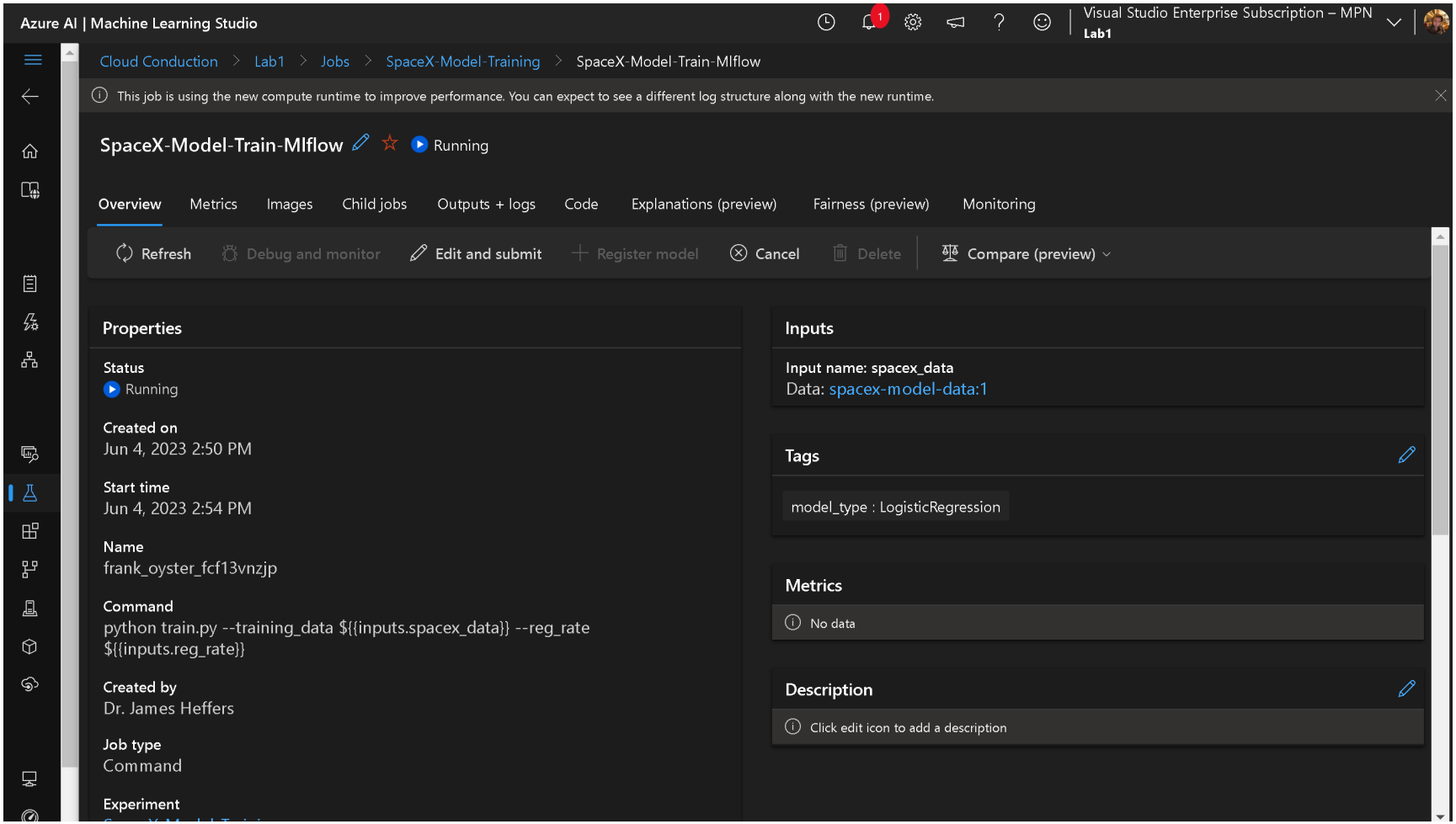


Modeling

- The full details for the several different models can be found in the Hyperparameter tuning notebook.
- Since all of the models are roughly the same accuracy, we will just stick with a logistic regression model. We create and run an MLflow script (contained in the MLflow notebook).
- After running the script we will register the model in Azure. Being satisfied with out model, we deploy it as a real-time inference endpoint.

Modeling

Our registered model!

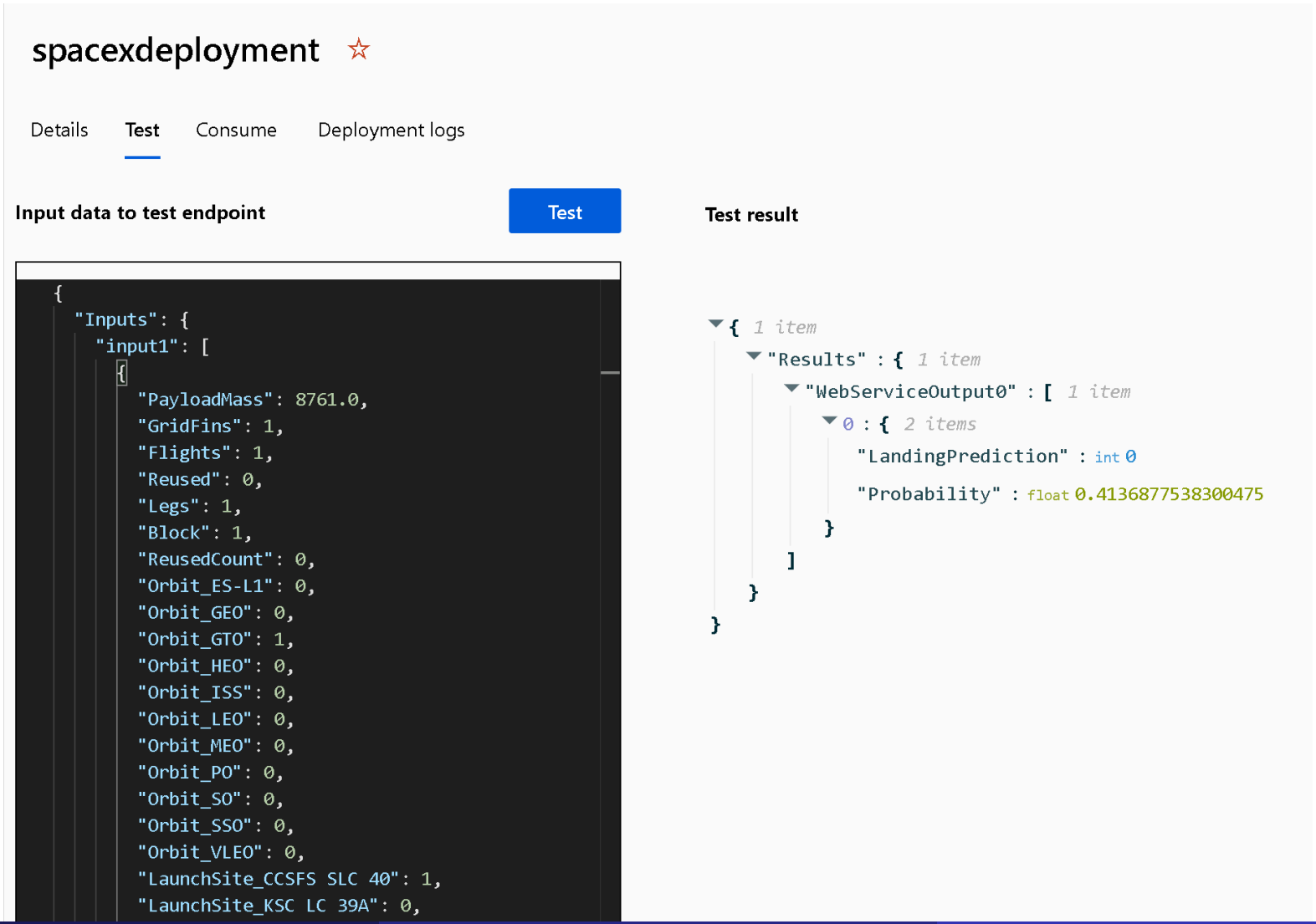


Modeling

- Our model is deployed for consumption, but let's finish by performing a sanity test.
- From our insights, our gut instincts tell us that a rocket with a heavy payload going into high orbit (such as GTO) will be unlikely to have a successful landing. A rocket with a light payload going into a low orbit (say, LEO), should have a good chance of successfully being reused.
- Recall that a label of 0 means failure to land, and a label of 1 means a successful landing.

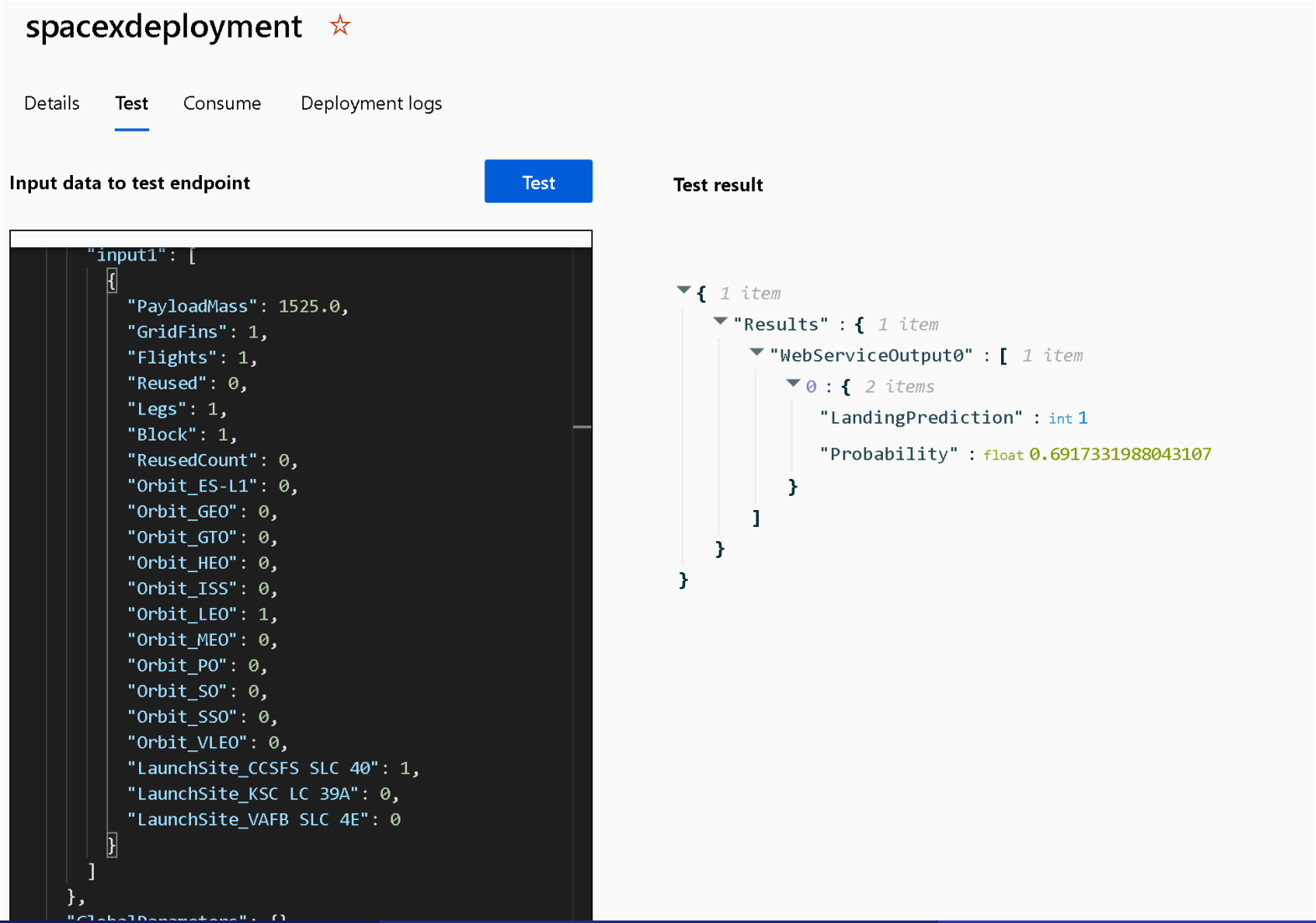
Modeling

We see the model predicts the following rocket (heavy payload, high orbit) will fail to land successfully.



Modeling

We see the model predicts the following rocket (light payload, low orbit) will indeed land successfully.



The End!

Thanks for checking out my project! If you did not get this PDF from my GitHub, then the link to the corresponding repository containing all the code used is below:

GitHub: https://github.com/TheProfessor712/IBMSpaceXProject712