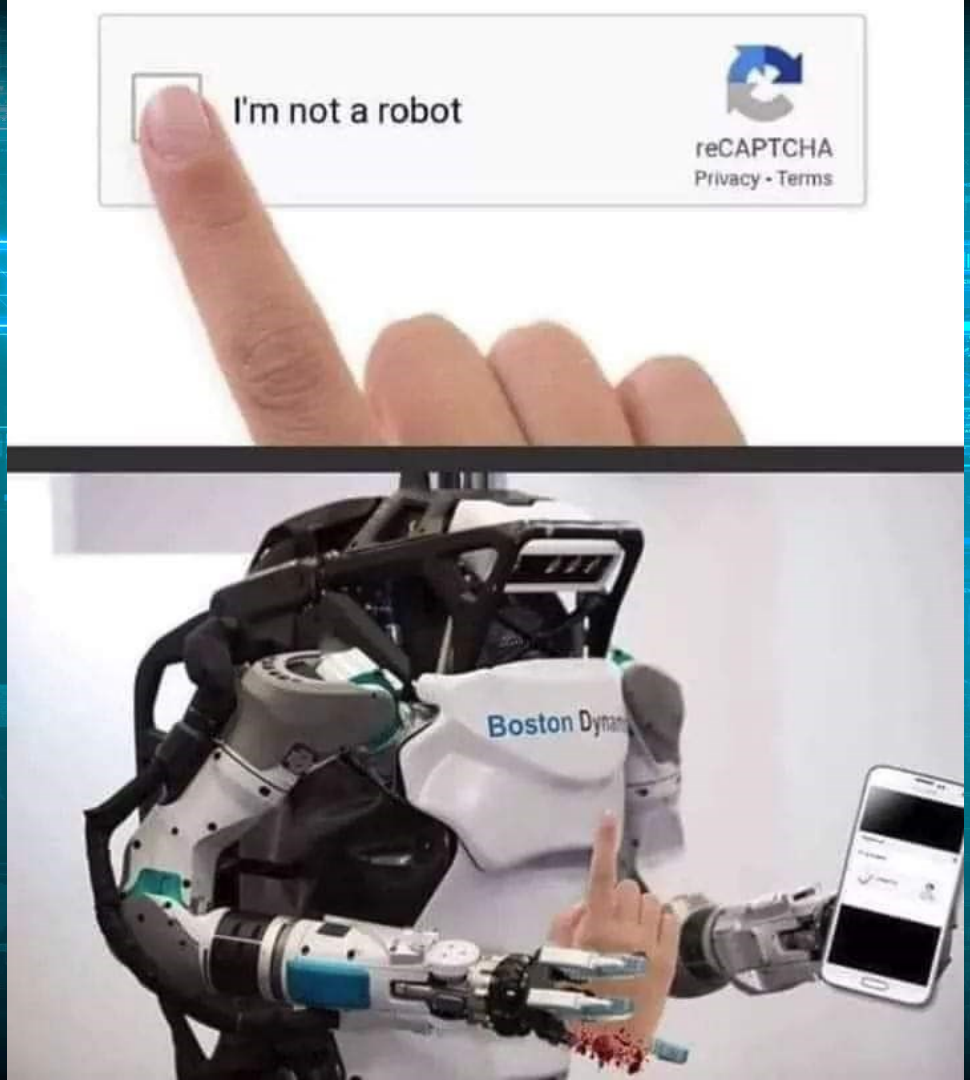
Come with me
if you want to
learn ML...
Session will start at
10:00 CST / 16:00 UTC

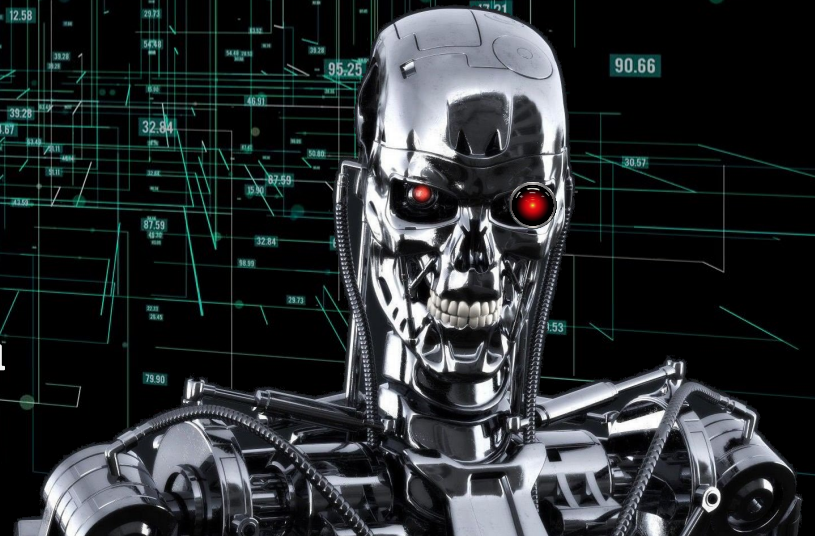


Session will start at
10:00 CST / 16:00 UTC

A Lunar Cat production...

From the producer of...

Machines Are Learning. You Should Learn ML Too: An Azure Story



Andre' Melancia

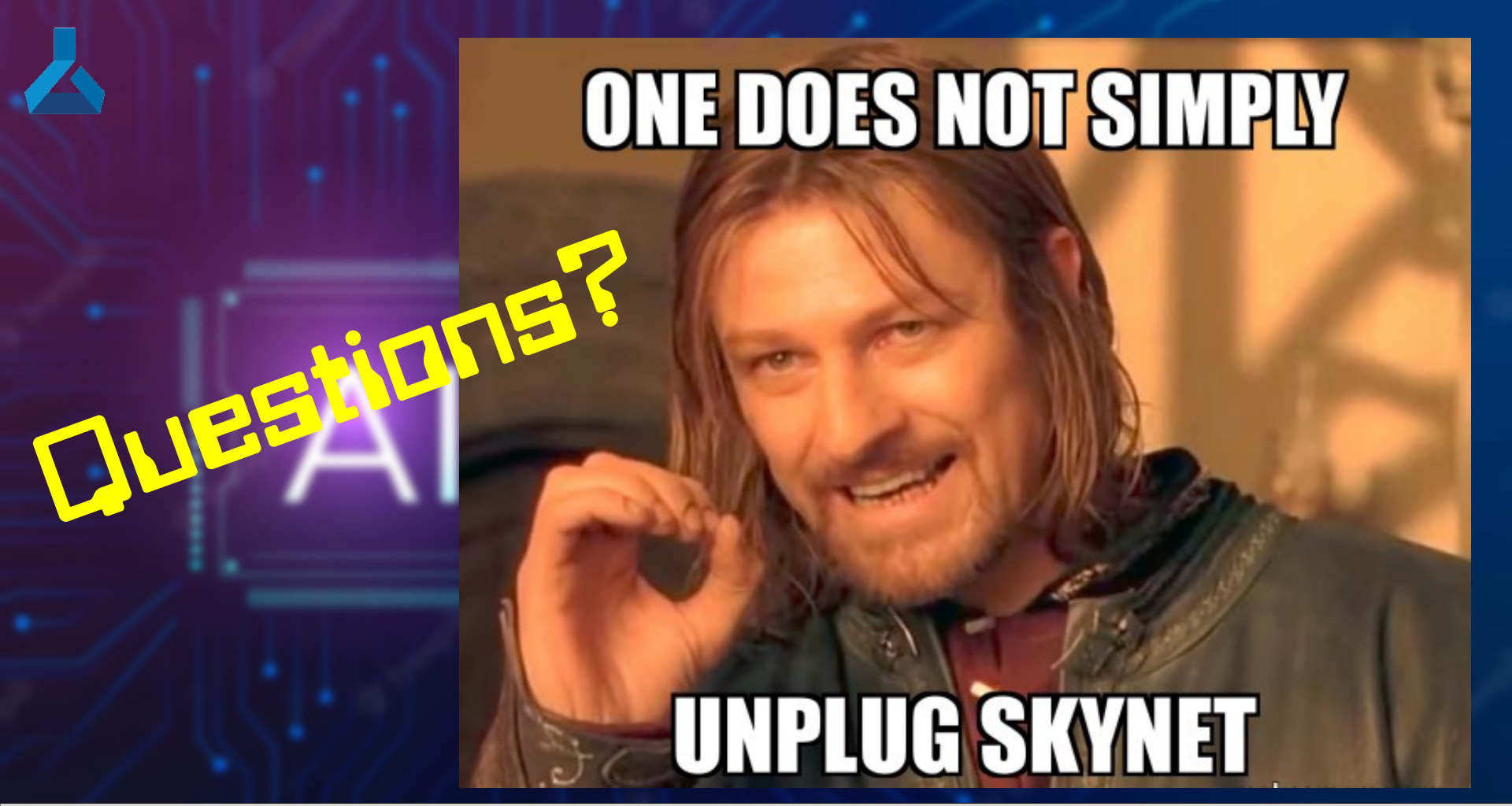
SQL Saturday #1028 Minnesota

2022-12-03

3 December
International Day of Persons with Disabilities

Accessible
United Nations

Be informed! Get involved!
www.un.org/disabilities
enable@un.org

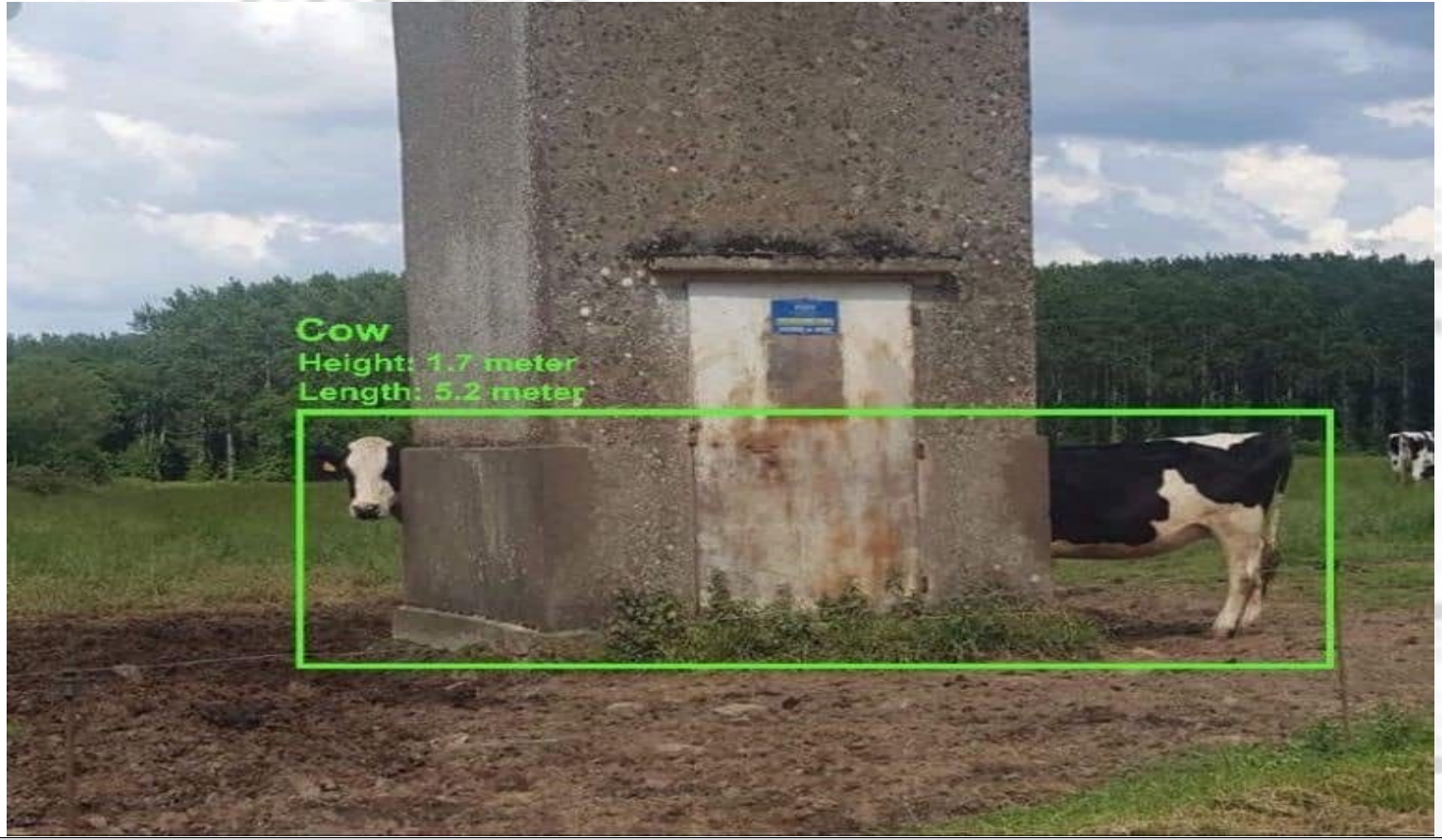


Questions?

Types of Algorithms

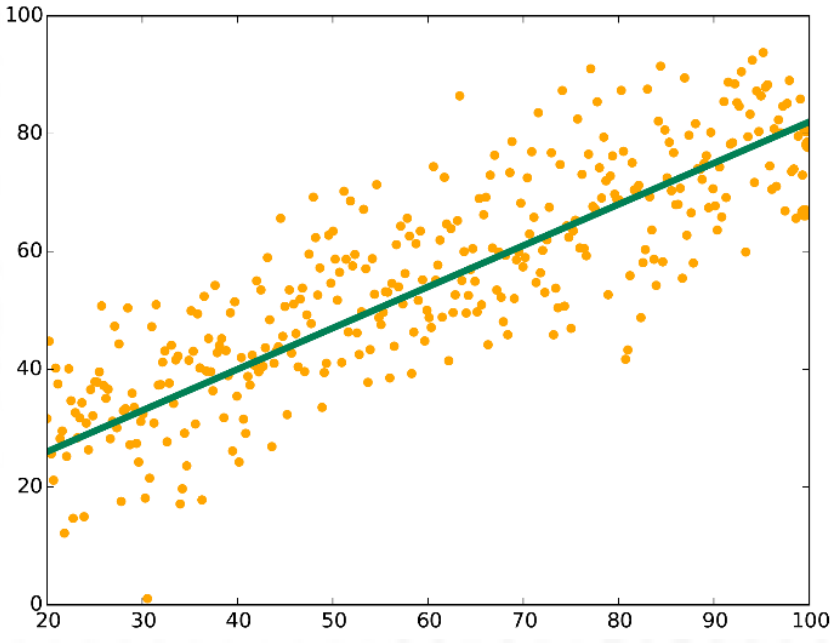
- Supervised (YOU provide the result)
  - Linear regression
  - Classification
- Unsupervised (YOU don't provide the result)
  - Clustering
  - Imaging (e.g. Neural networks)

Advanced bovine statistics

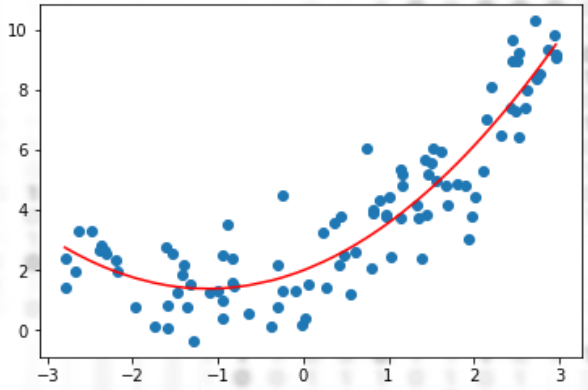


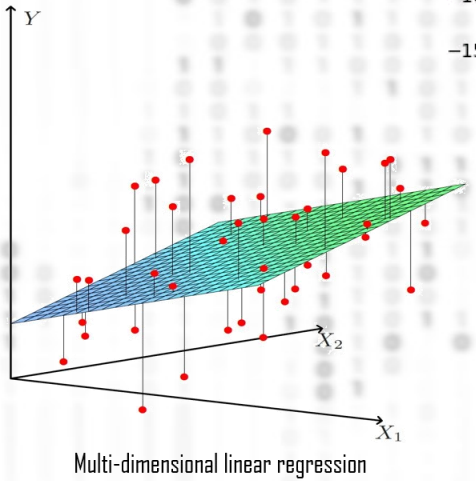
Regression Algorithms

- Learn by finding a linear pattern in data
  - https://en.wikipedia.org/wiki/Linear_regression
- Predict numeric values (linear or not)
  - Trends
  - Weather
  - Economics
  - Biology
  - [...]

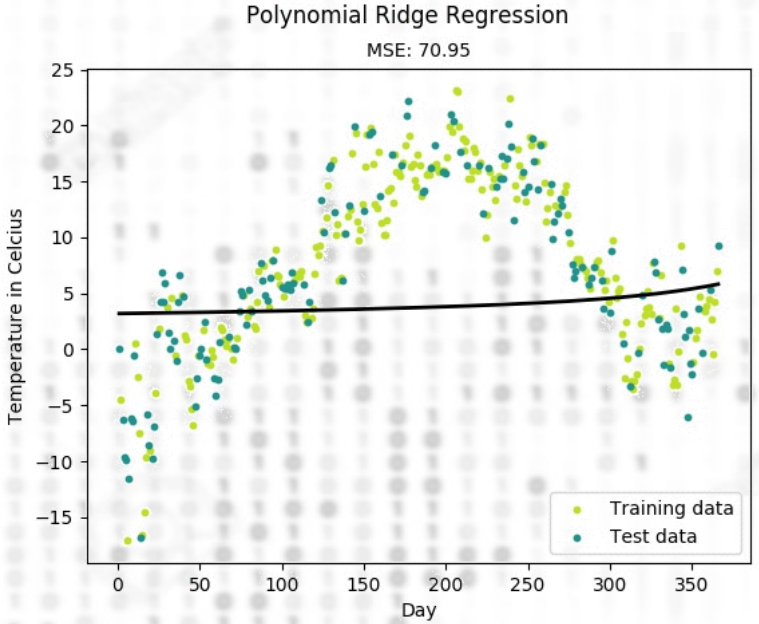


Algorithm: Regression



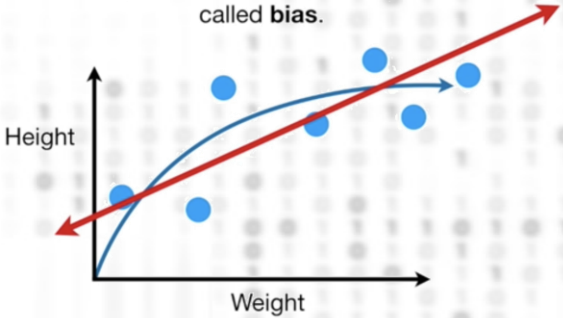


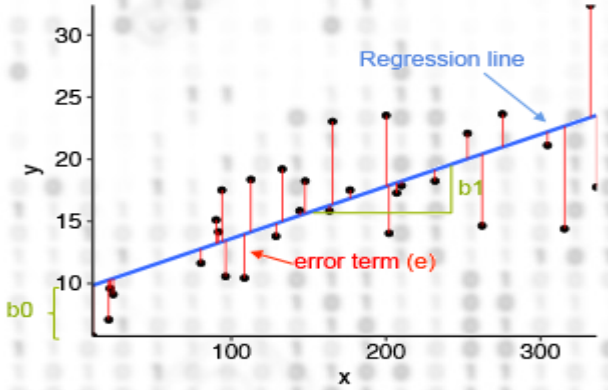
Multi-dimensional linear regression

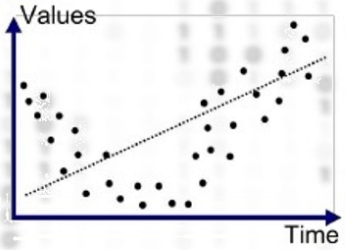


Algorithm: Regression

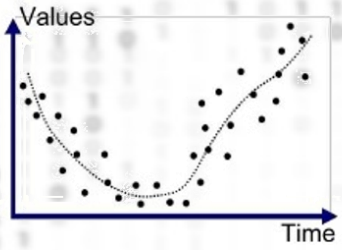
The inability for a machine learning method (like linear regression) to capture the true relationship is called **bias**.



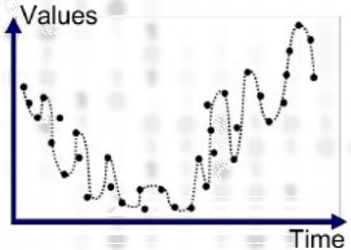




Underfitted

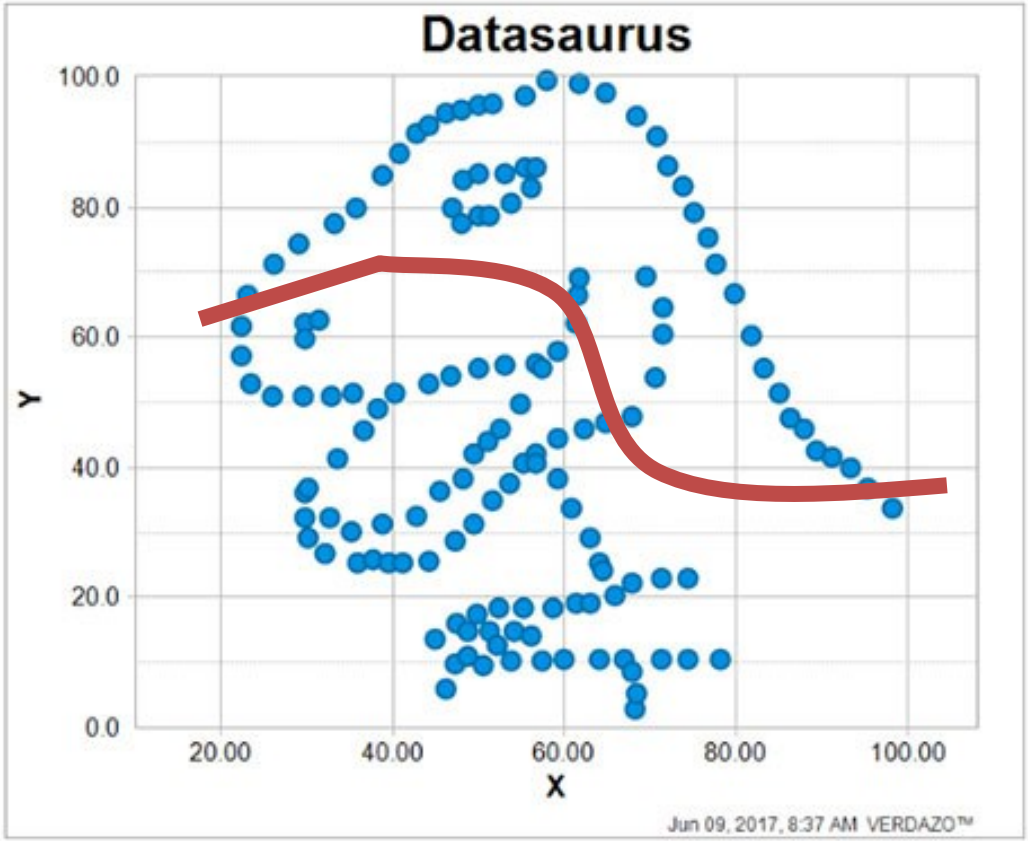


Good Fit/Robust



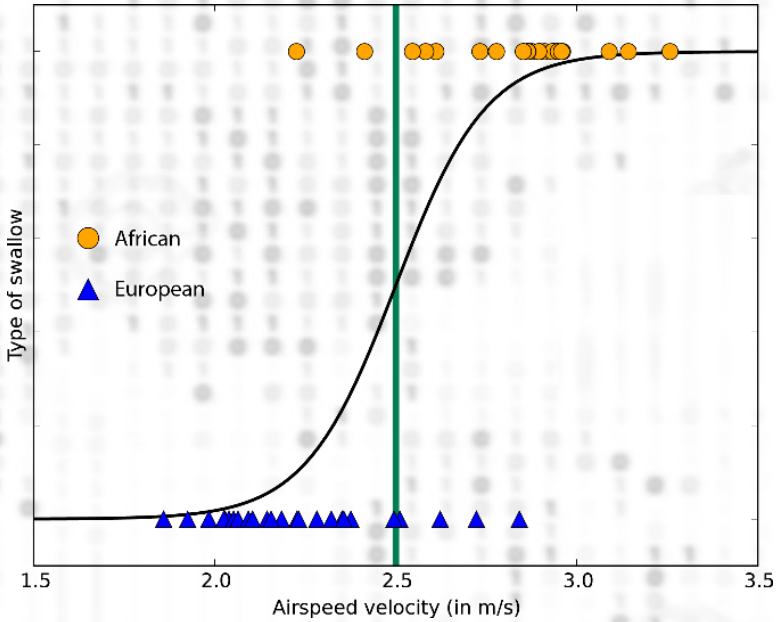
Overfitted

Beware of the statistics monsters...



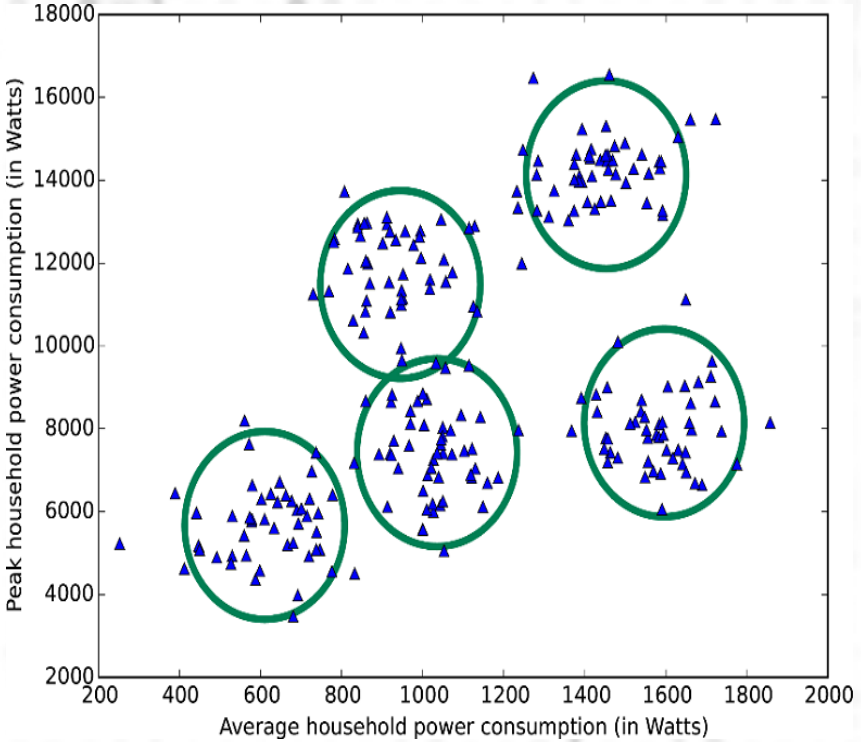
Classification Algorithms

- Classify in 2-class or multi-class models
- Logistic Regression (2 class)
  - https://en.wikipedia.org/wiki/Logistic_regression
  - Learn by finding a binary pattern in data
  - Either the positive or the negative (or A and B)
  - Returns a numeric value (e.g. percentage, where 0% is most likely false and 100% is most likely true)



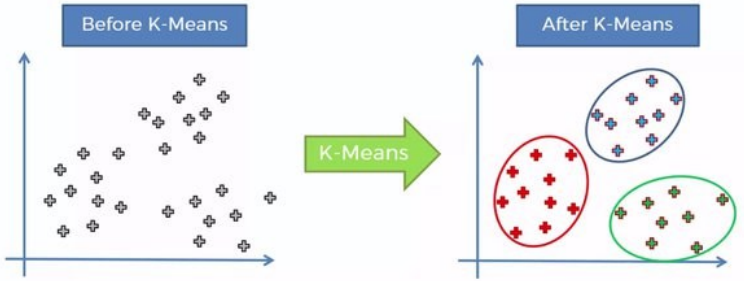
Clustering Algorithm

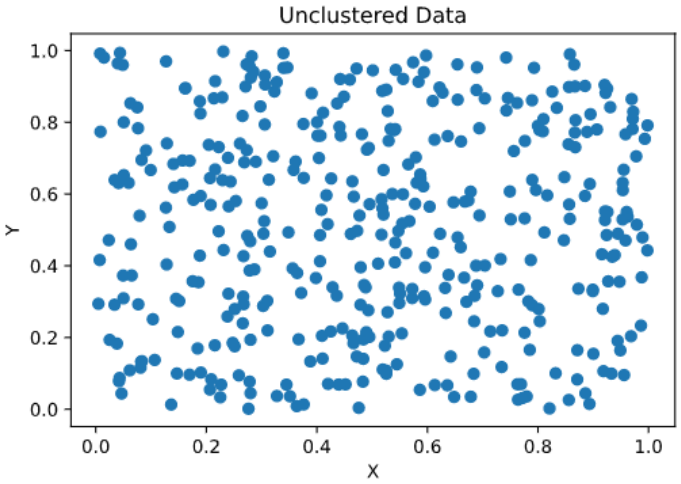
- Multiple algorithms for clustering
- k-means, etc.
  - Find grouping patterns in data
  - Shopping trends
  - Social network analysis
  - Crime analysis (fraud detection)

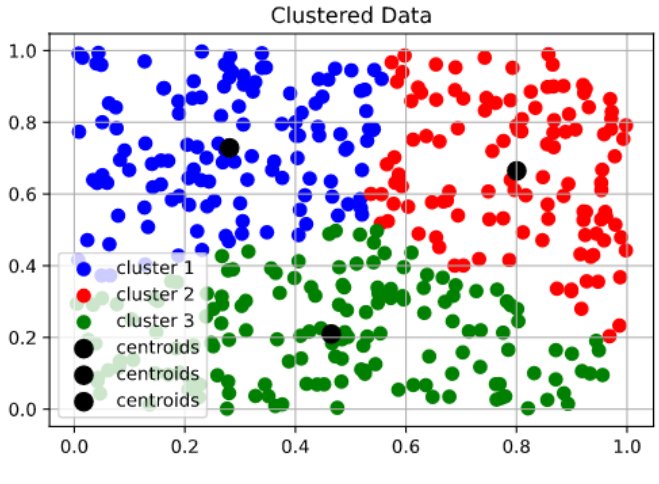


- https://en.wikipedia.org/wiki/K-means_clustering

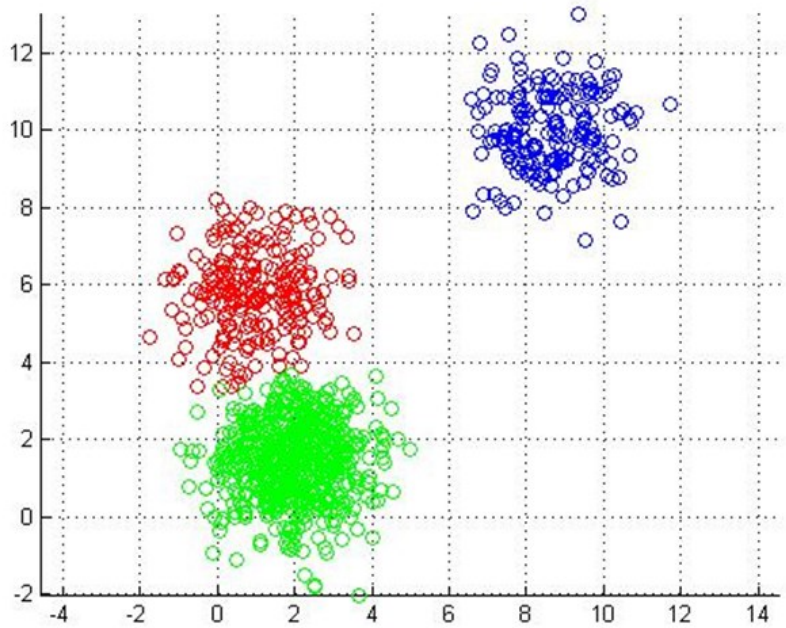
Clustering with K-means

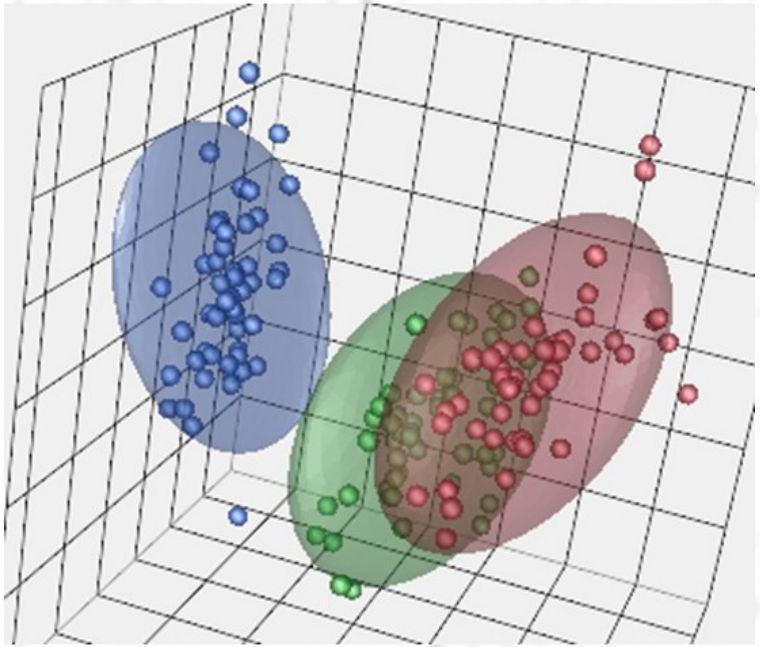






Algorithm: Clustering in N-dimensions



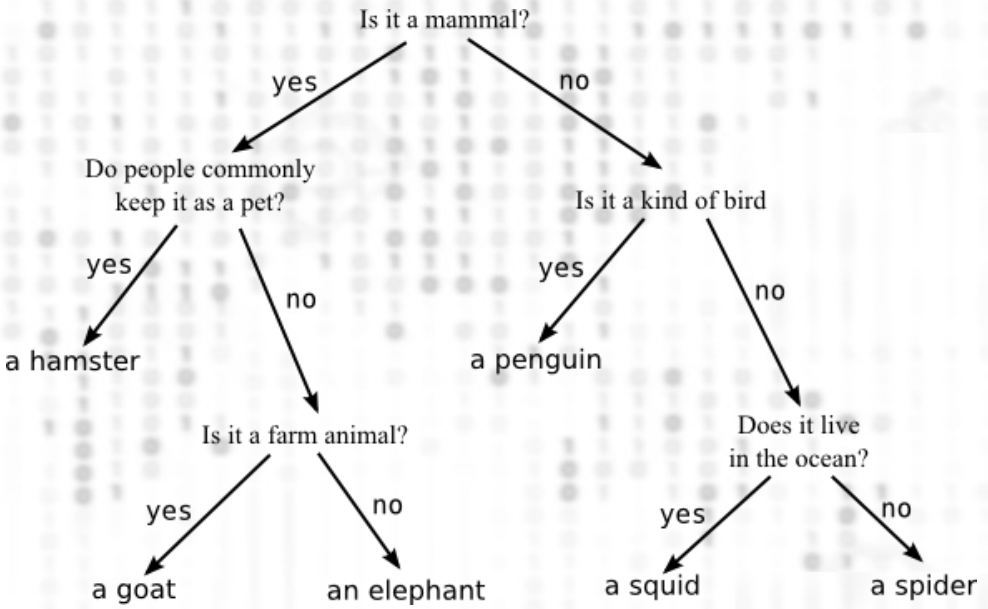


Explanation + code : https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

Algorithm techniques: Decision tree

- E.g.Multi-class model
- Both for numerical and categorical data.
- Conditions explained by simple Boolean logic.
- Easy to expand the tree if more detail is needed

- https://en.wikipedia.org/wiki/Decision_tree_learning

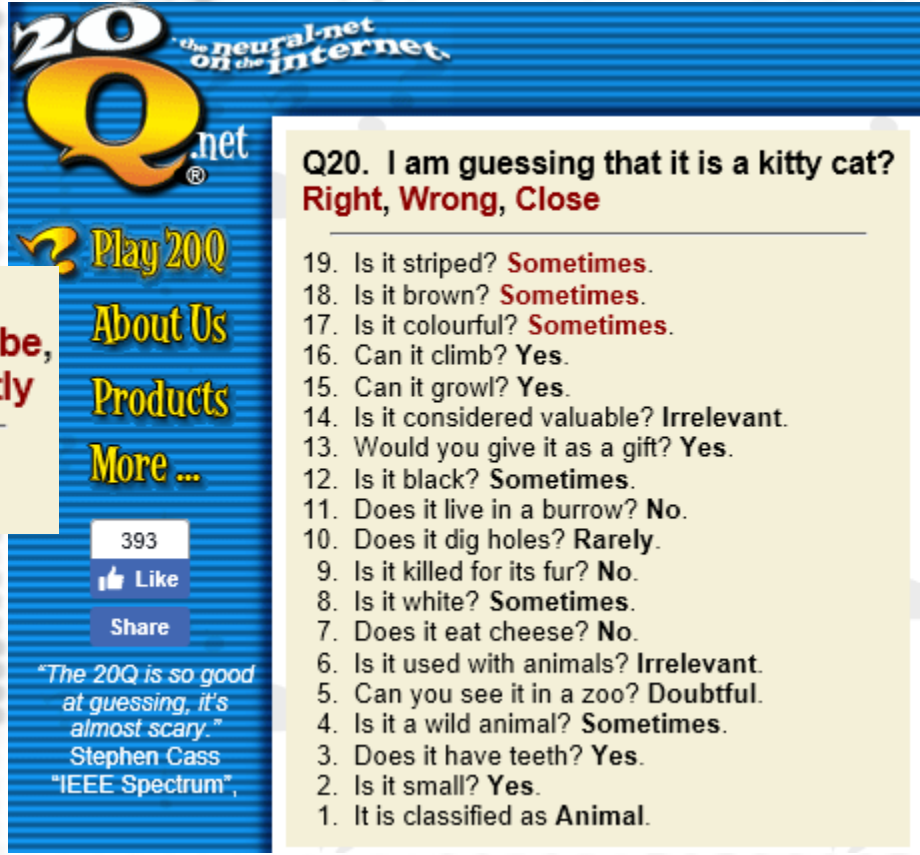


Decision tree in the wild (since the 1990s)

- http://20Q.NET

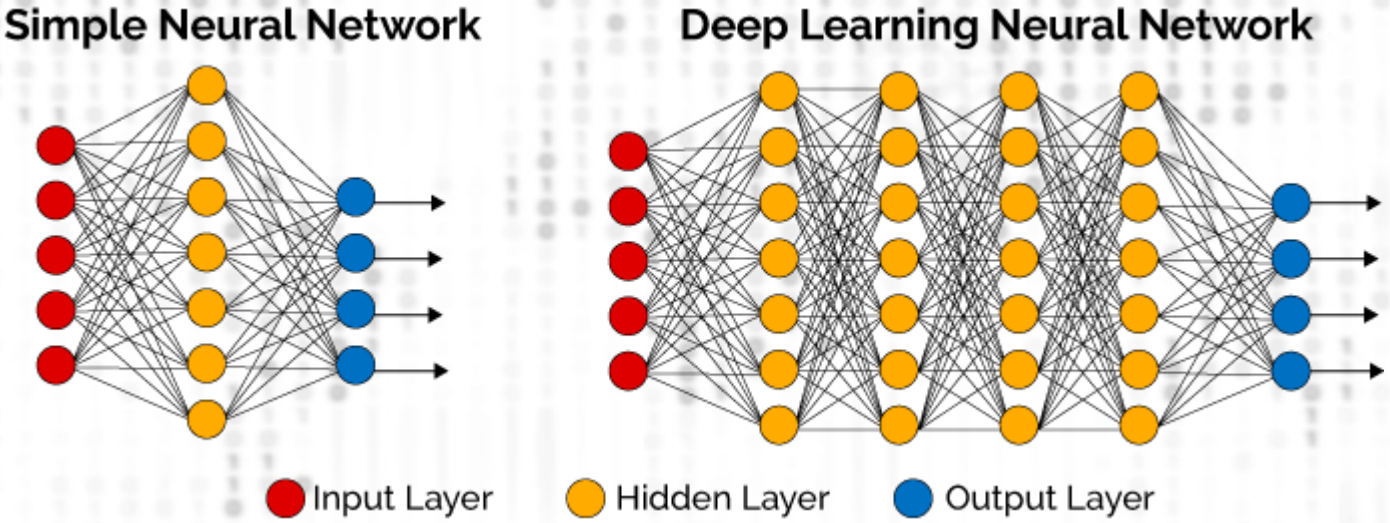
Q3. Will it eat almost anything?
Yes , No , Unknown, Irrelevant, Sometimes, Maybe, Probably, Doubtful, Usually, Depends, Rarely, Partly

2. Can it scratch? **Yes**.
1. It is classified as **Animal** .



Algorithm techniques: Neural Networks

- E.g.Imaging, Deep Learning
- Slowest but most accurate
- https://en.wikipedia.org/wiki/Artificial_neural_network
- https://en.wikipedia.org/wiki/Deep_learning



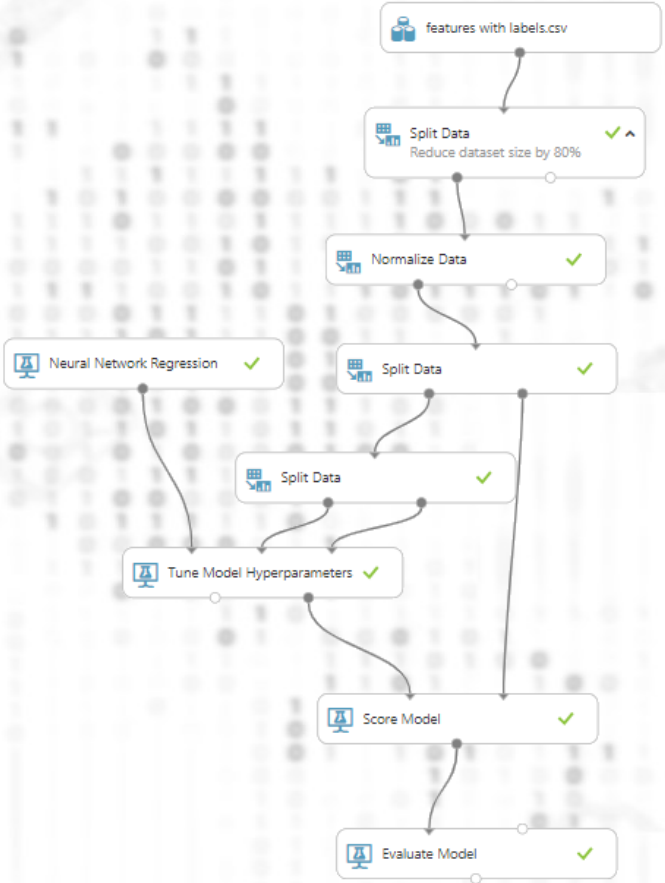
Not all "Data Science" is Data Science...

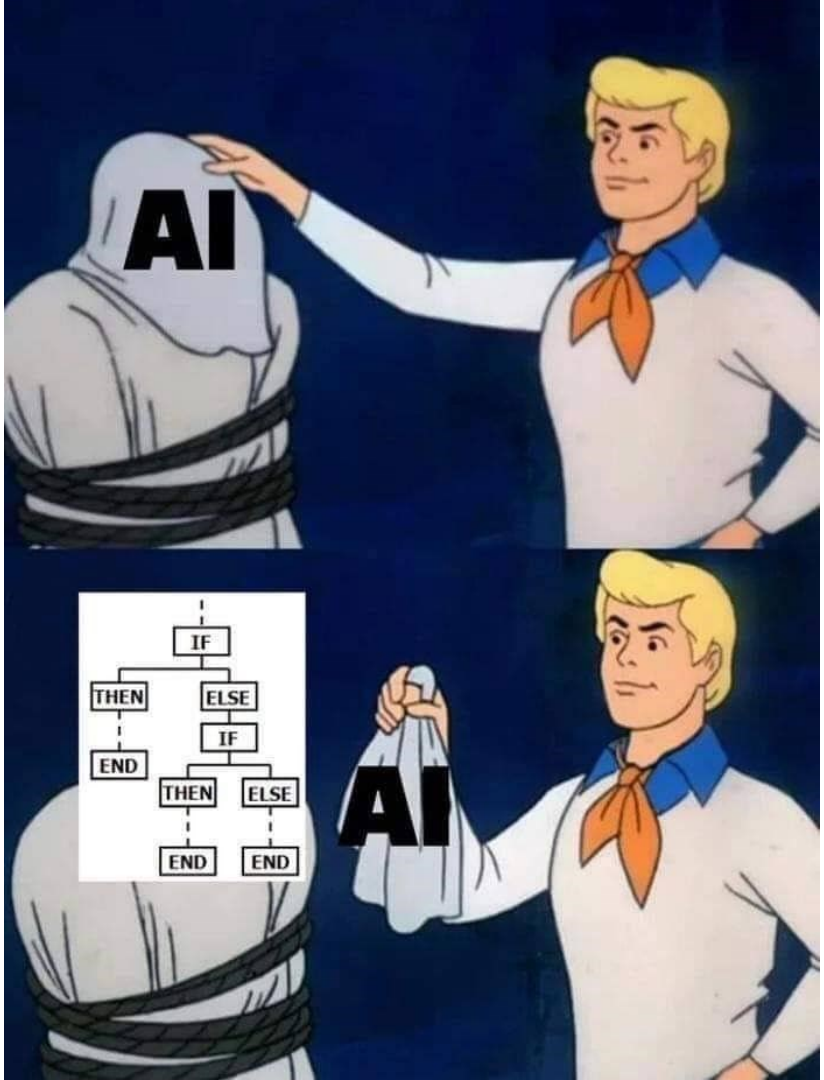
For all of you naturally skeptical science lovers:
Does this qualify as a sandwich?

I was told workflow was here. Could I see it, please?

Machine Learning typical workflow:

1. Get datasets
2. Clean, prep, feature engineering
3. Training (Choosing Algorithm and Hyper-Parameters)
   - Regression, Classification (2-class, multi-class), Clustering, Anomaly Detection, etc.
   - Deep Learning, Neural Networks, etc.
4. Scoring/Testing
5. Evaluating
6. Deploying predictive webservices (inference)
7. I'll retrain back!




AI
IF
THEN
ELSE
IF
END
THEN
ELSE
END
END
AI

Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution.
Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

Microsoft Azure

What do you want to do?

Extract information from text

Text Analytics

Derives high-quality information from text

Answers questions like: What info is in this text?

- Latent Dirichlet Allocation ← Unsupervised topic modeling, group texts that are similar
- Extract N-Gram Features from Text ← Creates a dictionary of n-grams from a column of free text
- Feature Hashing ← Converts text data to integer encoded features using the Vowpal Wabbit library
- Preprocess Text ← Performs cleaning operations on text, like removal of stop-words, case normalization
- Word2Vector ← Converts words to values for use in NLP tasks, like recommender, named entity recognition, machine translation

Predict between several categories

Multiclass Classification

Answers complex questions with multiple possible answers

Answers questions like: Is this A or B or C or D?

- Multiclass Logistic Regression ← Fast training times, linear model
- Multiclass Neural Network ← Accuracy, long training times
- Multiclass Decision Forest ← Accuracy, fast training times
- One-vs-All Multiclass ← Depends on the two-class classifier
- One-vs-One Multiclass ← Depends on binary classifier, less sensitive to an imbalanced dataset with larger complexity
- Multiclass Boosted Decision Tree ← Non-parametric, fast training times and scalable

Generate recommendations

Recommenders

Predicts what someone will be interested in

Answers the question: What will they be interested in?

- Use the Train Wide & Deep Recommender module ← Hybrid recommender, both collaborative filtering and content-based approach
- SVD Recommender ← Collaborative filtering, better performance with lower cost by reducing dimensionality

Predict between two categories

Two-Class Classification

Answers simple two-choice questions, like yes or no, true or false

Answers questions like: Is this A or B?

- Two-Class Support Vector Machine ← Under 100 features, linear model
- Two-Class Averaged Perceptron ← Fast training, linear model
- Two-Class Decision Forest ← Accurate, fast training
- Two-Class Logistic Regression ← Fast training, linear model
- Two-Class Boosted Decision Tree ← Accurate, fast training, large memory footprint
- Two-Class Neural Network ← Accurate, long training times

Predict values

Regression

Makes forecasts by estimating the relationship between values

Answers questions like: How much or how many?

- Fast Forest Quantile Regression ← Predicts a distribution
- Poisson Regression ← Predicts event counts
- Linear Regression ← Fast training, linear model
- Bayesian Linear Regression ← Linear model, small data sets
- Decision Forest Regression ← Accurate, fast training times
- Neural Network Regression ← Accurate, long training times
- Boosted Decision Tree Regression ← Accurate, fast training times, large memory footprint

Discover structure

Clustering

Separates similar data points into intuitive groups

Answers questions like: How is this organized?

- K-Means ← Unsupervised learning

Classify images

Image Classification

Classifies images with popular networks

Answers questions like: What does this image represent?

- ResNet ← Modern deep learning neural network
- DenseNet ← Modern deep learning neural network

Find unusual occurrences

Anomaly Detection

Identifies and predicts rare or unusual data points

Answers the question: Is this weird?

- One Class SVM ← Under 100 features, aggressive boundary
- PCA-Based Anomaly Detection ← Fast training times

© 2021 Microsoft Corporation. All rights reserved.

Share this poster: aka.ms/mlcheatsheet

Humans...

Using Machine Learning in the Microsoft Universe doesn't require Data Science background!

Anyone can do it!

"This is the world now. Logged on, plugged in, all the time."

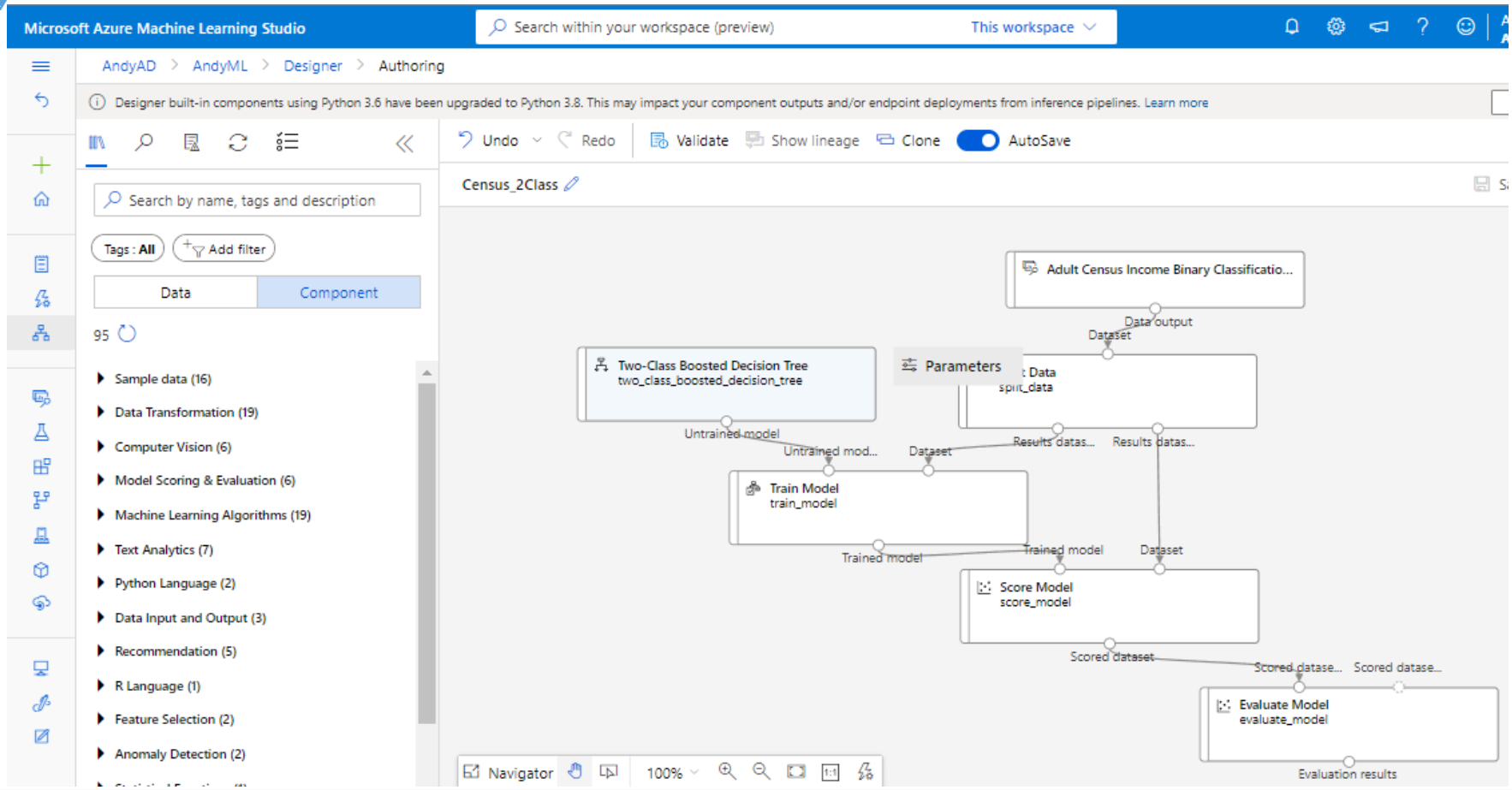
1. Azure Machine Learning (Service / Workspace, not "Classic")
   - SDK includes a lot of stuff (AutoML, HyperDrive, Model registry, SDK support for webservices in containers, etc.) – Supports Python, R (incomplete), MLFlow, etc.
2. Azure Databricks (Spark clusters, SQL, Python, R, Scala)
3. Azure HDInsight Spark
4. Azure HDInsight R Server
5. Azure Synapse Analytics (with SQL or Spark – SQL, Python, R, Scala, .Net)
6. SQL Server 2016 (SQL, .Net or R Services)
7. SQL Server 2017/2019/2022 (SQL, .Net, ML Services using R or Python)
8. Any platform with .Net Core and libraries for ML.Net
9. Azure Cognitive Services + Bots (managed webservices) + M365 Power Virtual Agents
10. Etc...

Magic is in the libraries, not the languages!

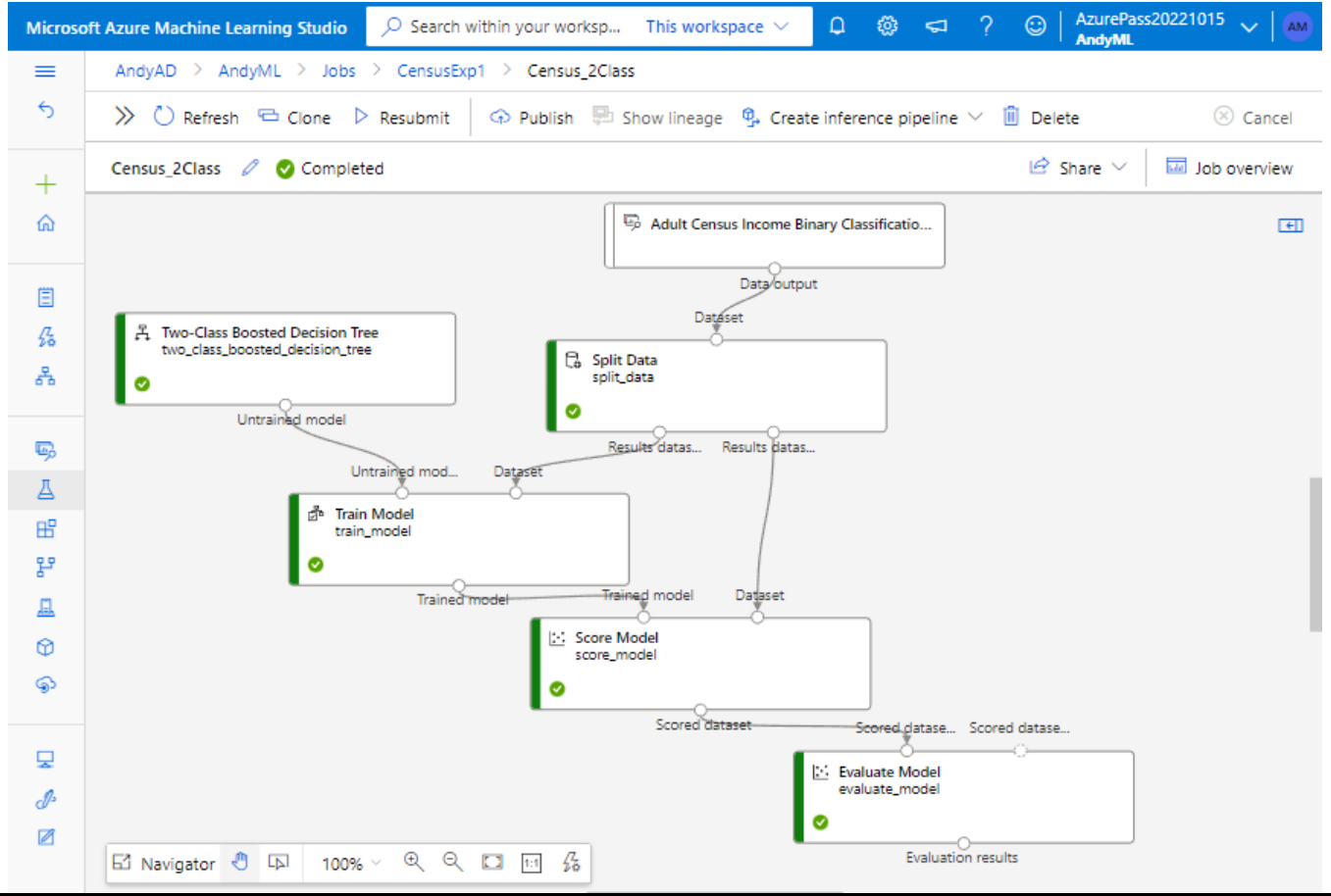
- Some of these include their own notebooks (Jupyter, Zeppelin, ODC, etc.) and you can use your tools like Azure Data Studio, Visual Studio Code, etc.

"Come with me if you want to DEMO..."

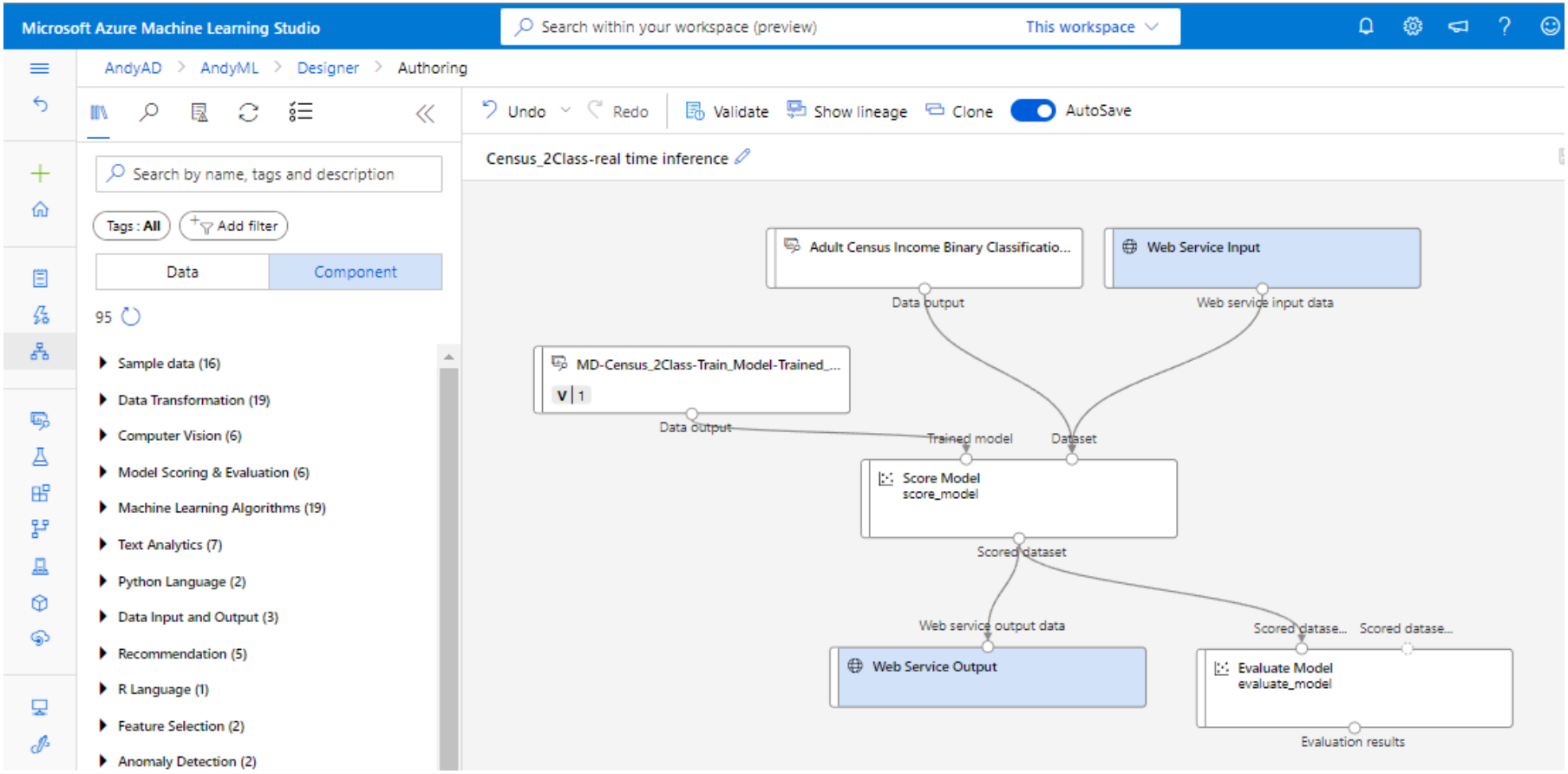
Bad demo? Here's screenshots!



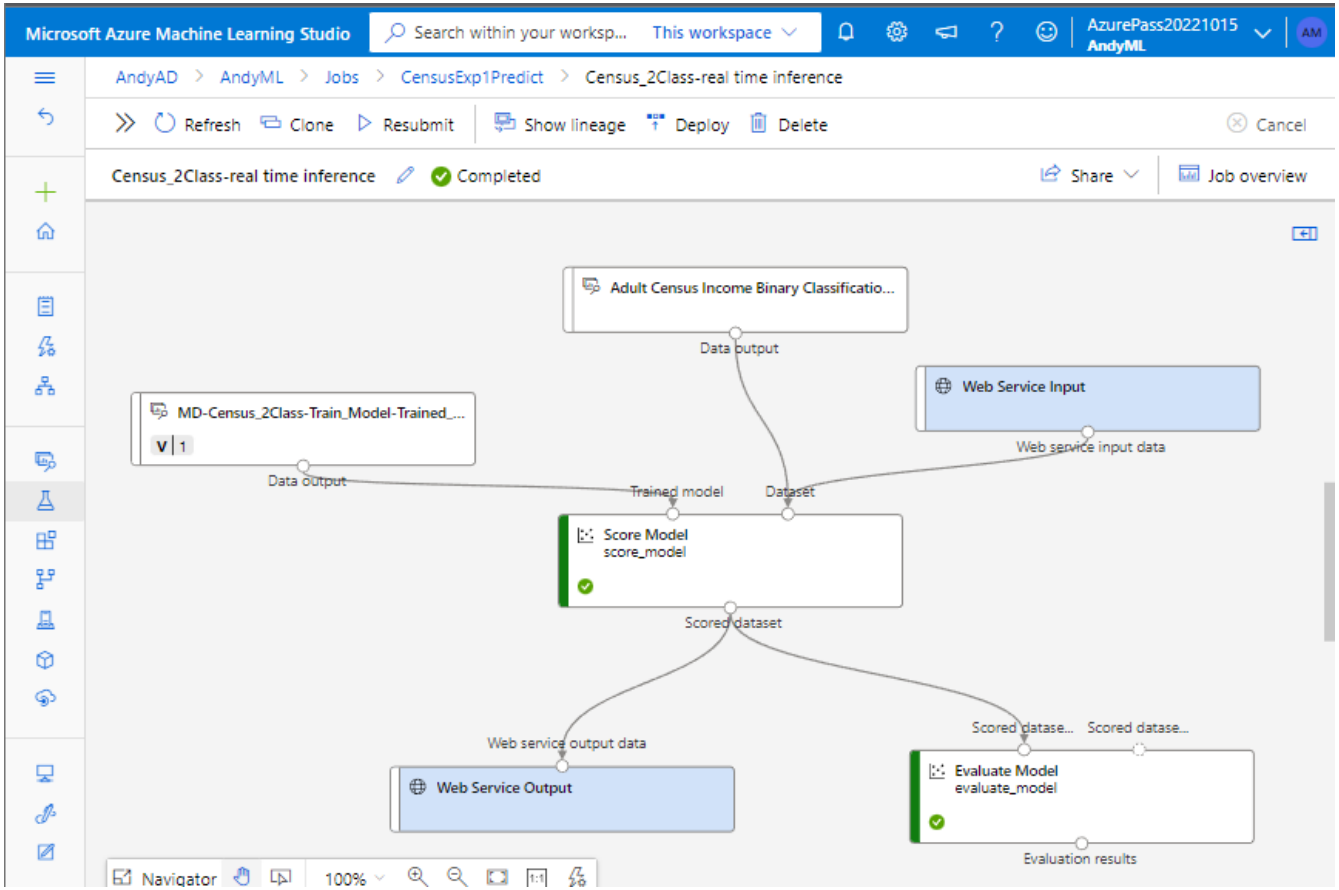
Bad demo? Here's screenshots!



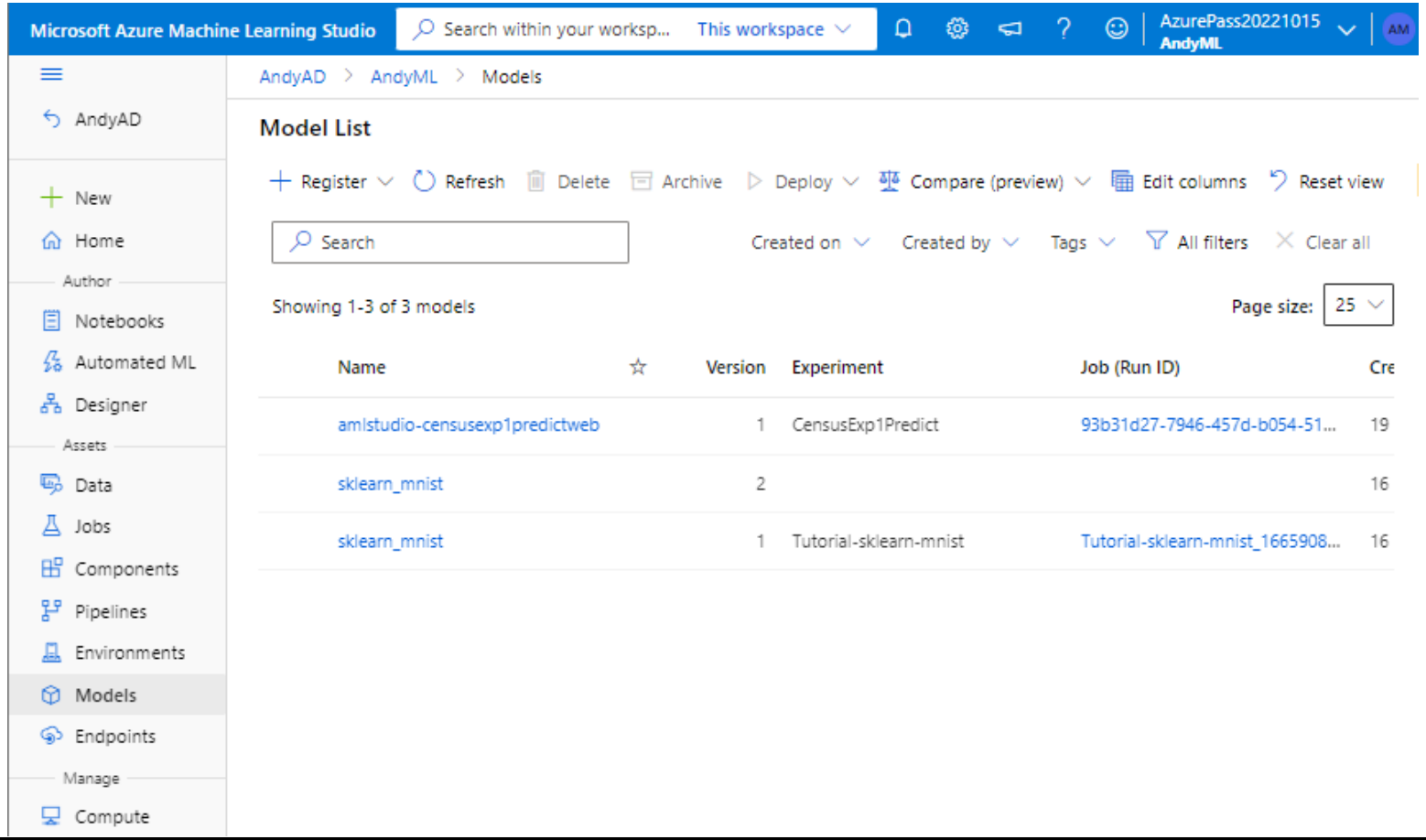
Bad demo? Here's screenshots!



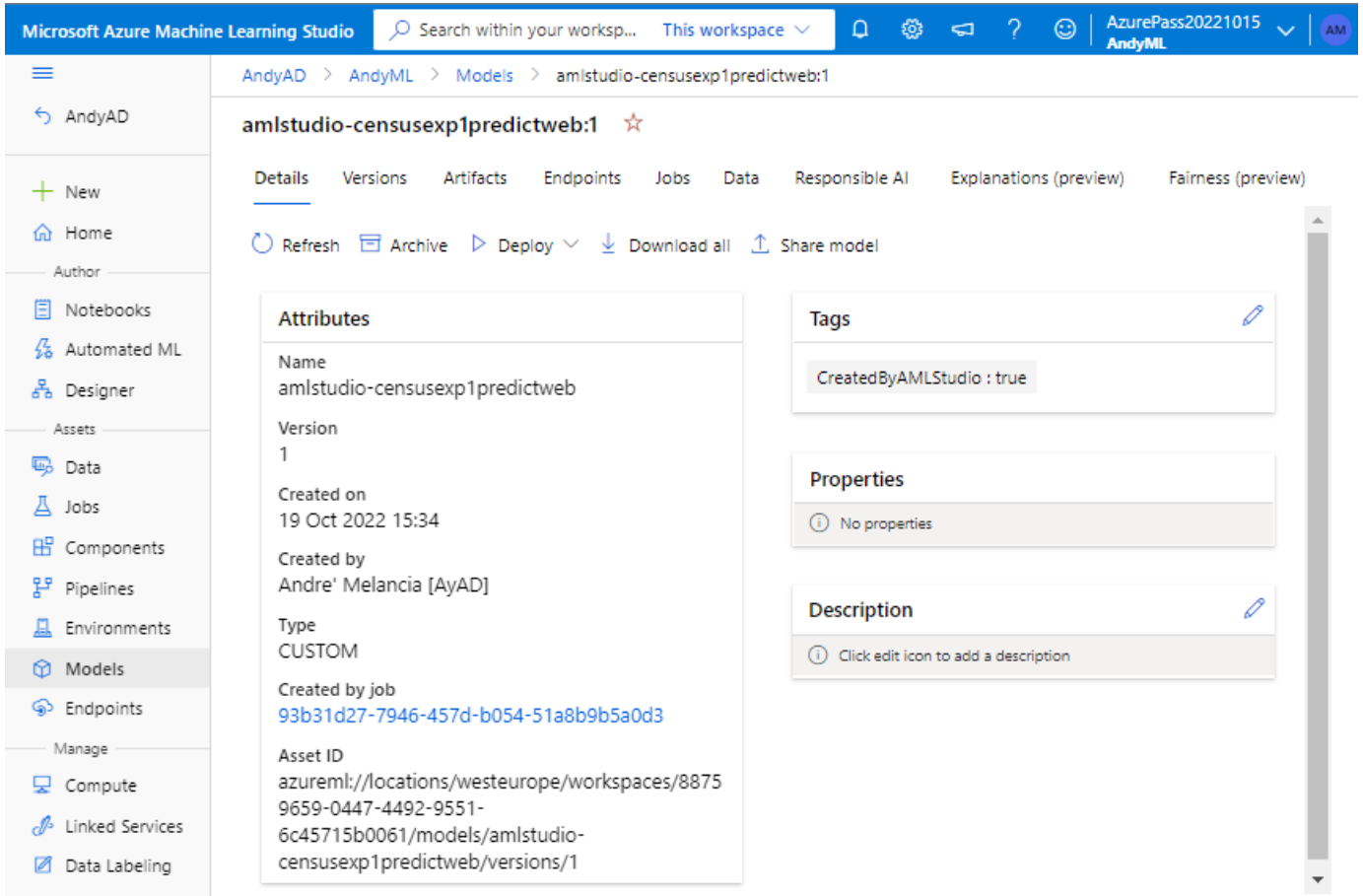
Bad demo? Here's screenshots!



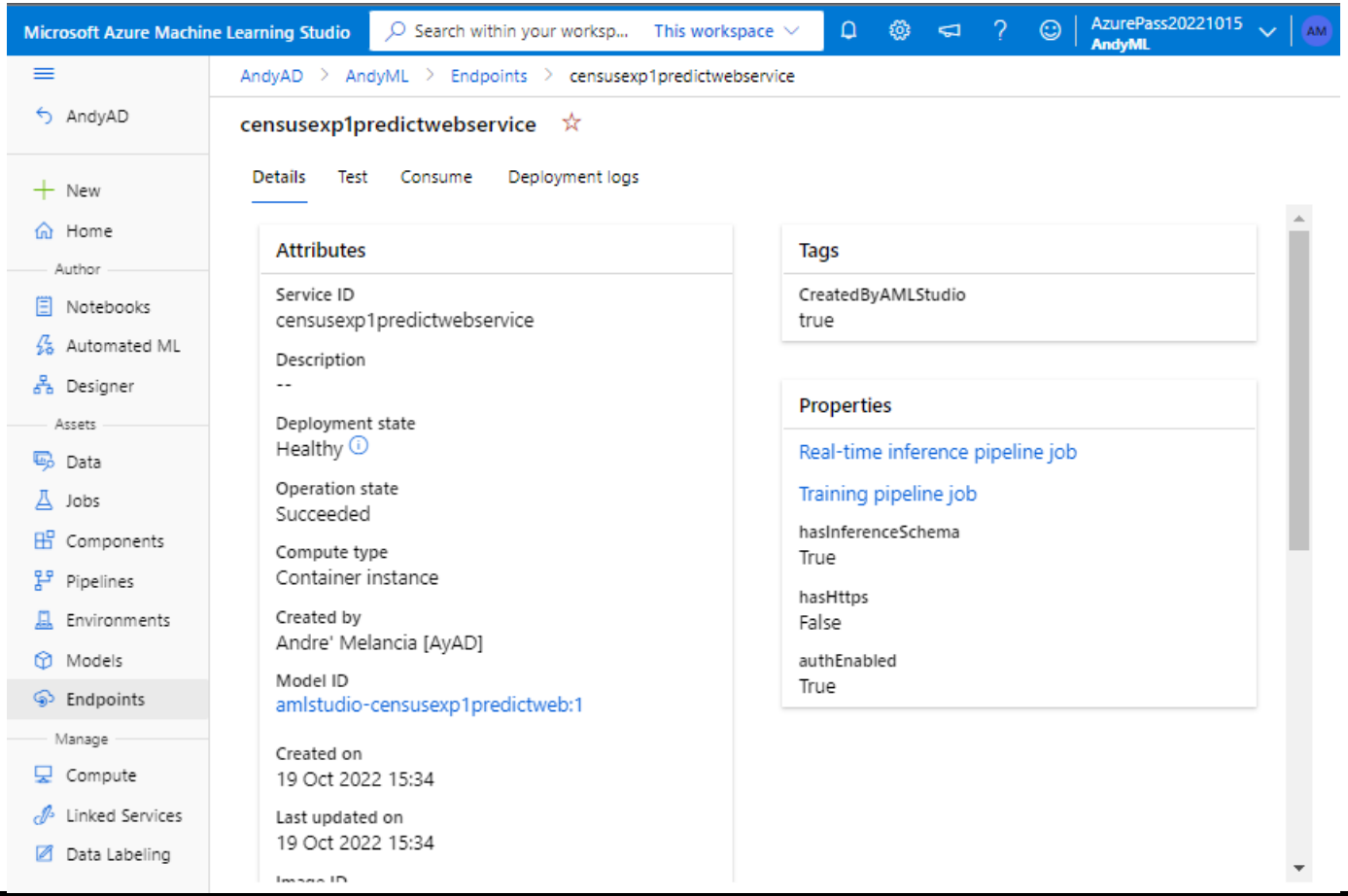
Bad demo? Here's screenshots!



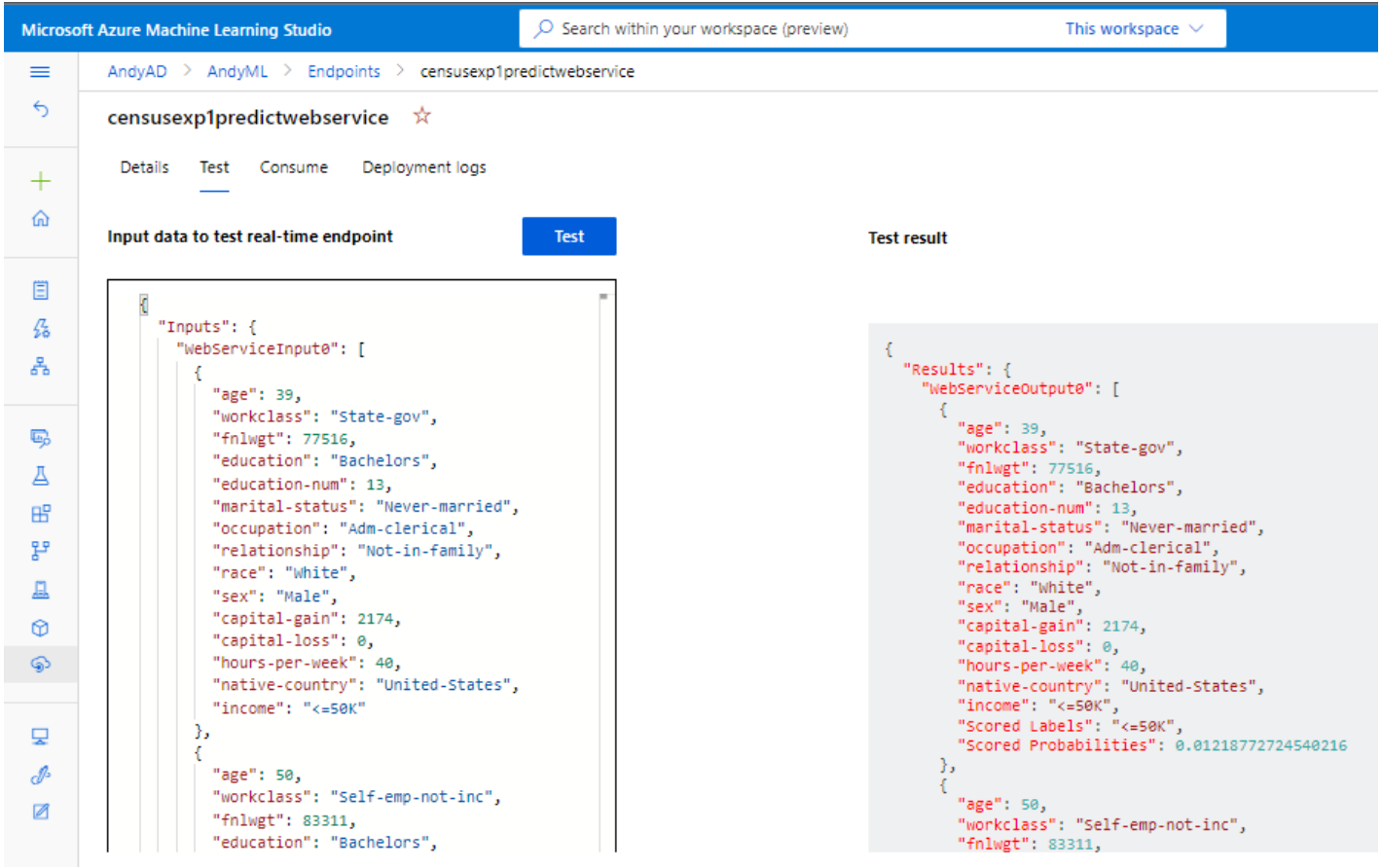
Bad demo? Here's screenshots!



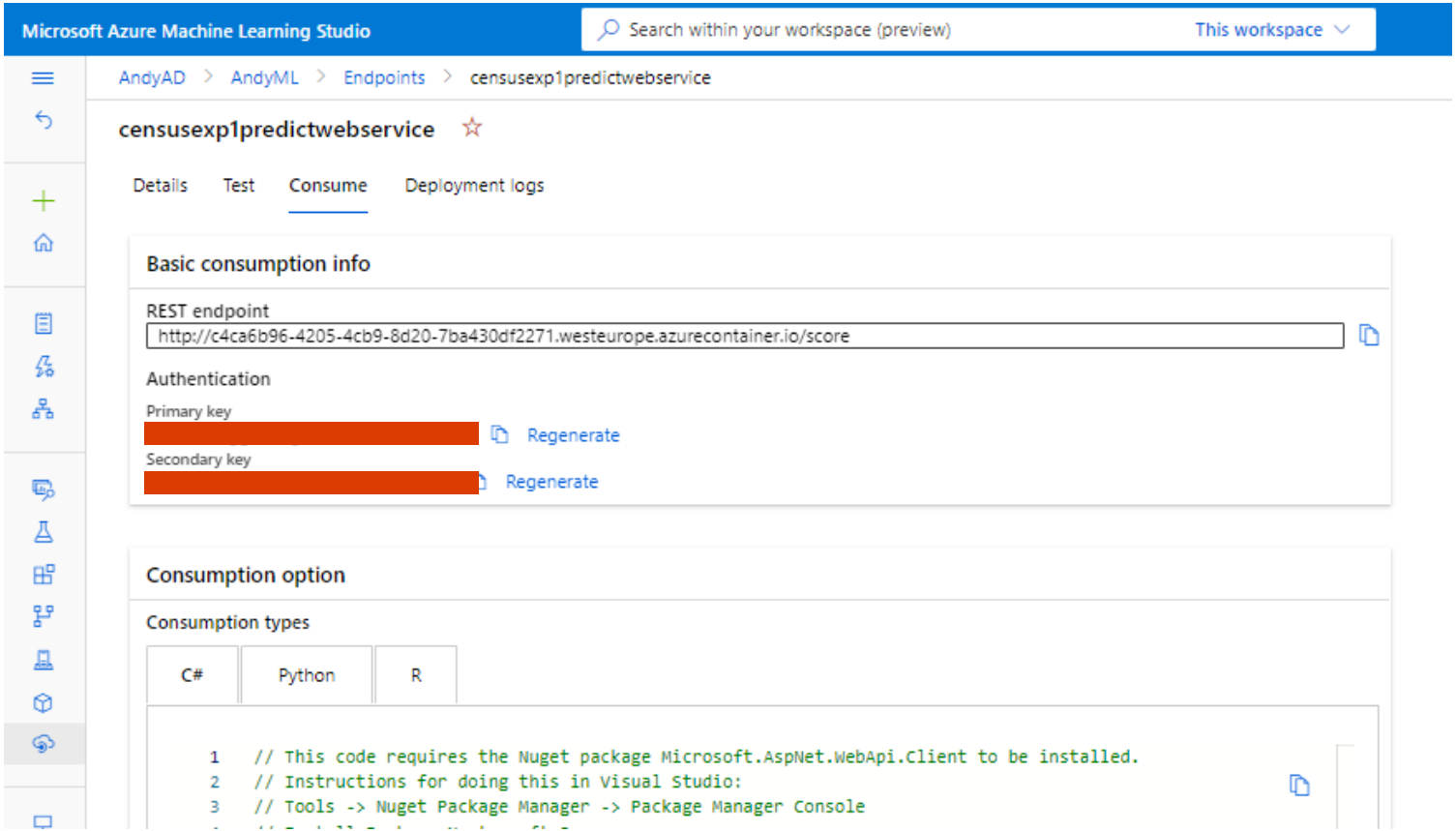
Bad demo? Here's screenshots!



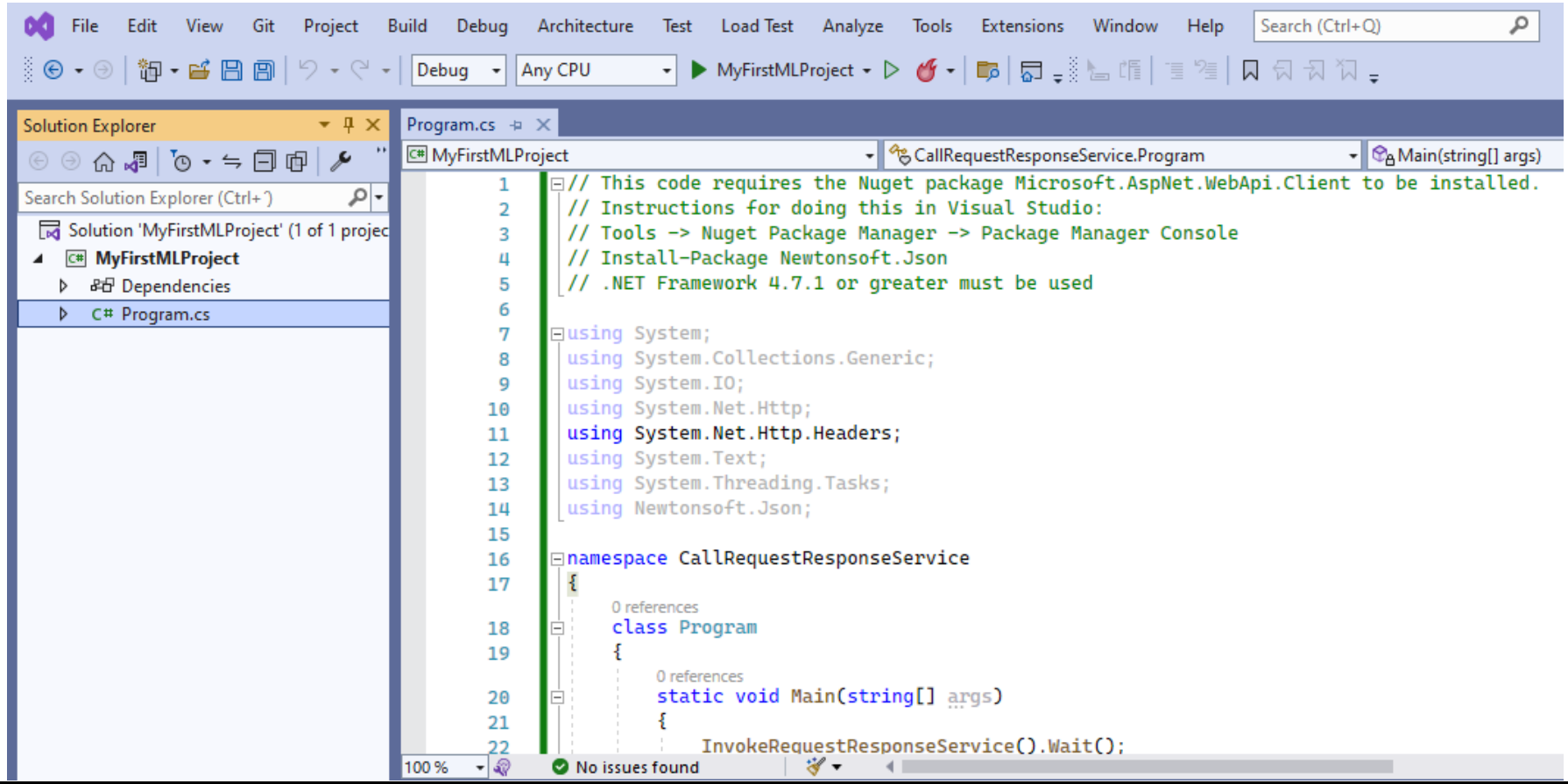
Bad demo? Here's screenshots!



Bad demo? Here's screenshots!



Bad demo? Here's screenshots!



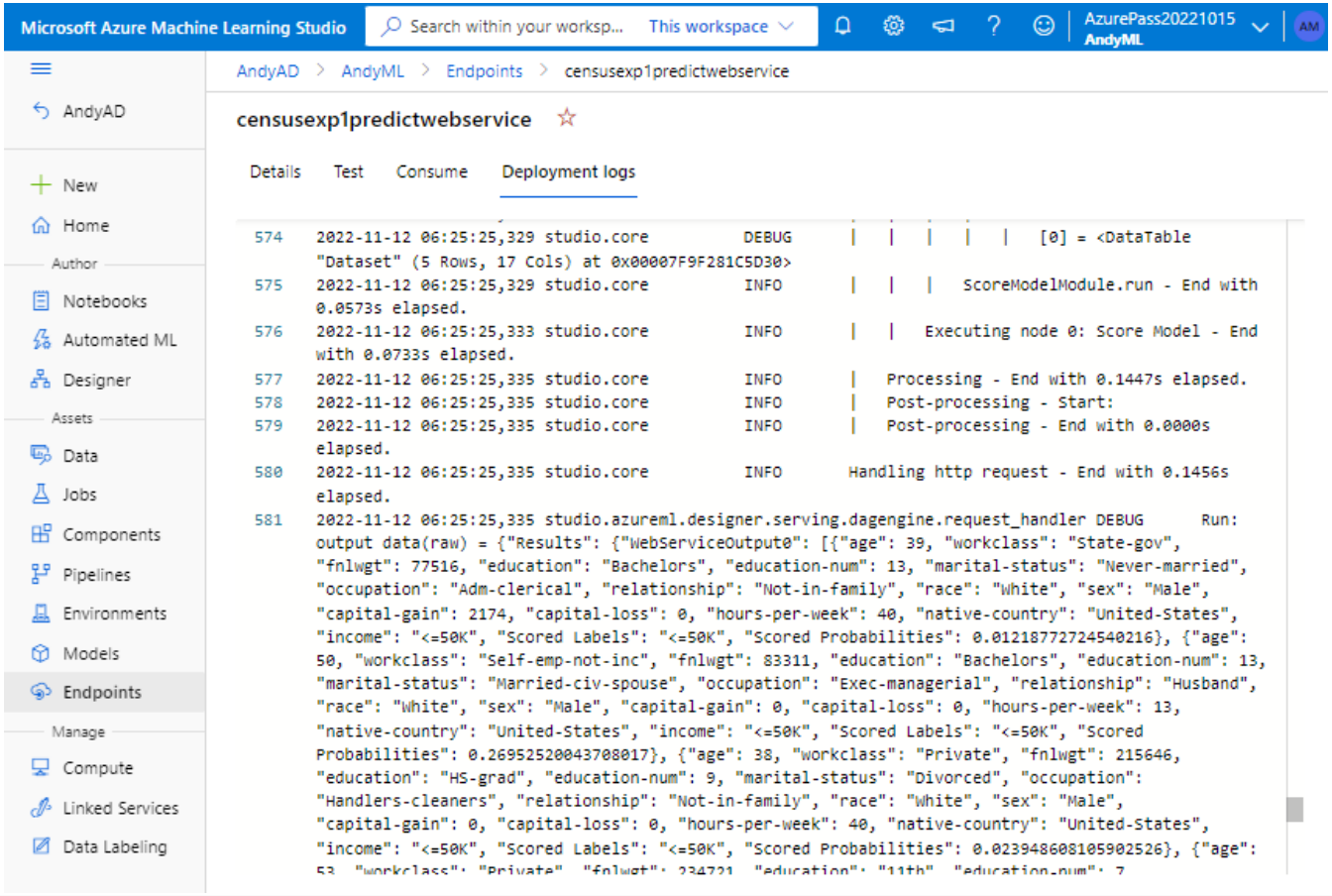
Bad demo? Here's screenshots!

Select Microsoft Visual Studio Debug Console


```
Result: {"Results": {"WebServiceOutput0": [{"age": 39, "workclass": "State-gov", "fnlwgt": 77516, "education": "Bachelor
s", "education-num": 13, "marital-status": "Never-married", "occupation": "Adm-clerical", "relationship": "Not-in-family
", "race": "White", "sex": "Male", "capital-gain": 2174, "capital-loss": 0, "hours-per-week": 40, "native-country": "Uni
ted-States", "income": "<=50K", "Scored Labels": "<=50K", "Scored Probabilities": 0.01218772724540216}, {"age": 50, "wor
kclass": "Self-emp-not-inc", "fnlwgt": 83311, "education": "Bachelors", "education-num": 13, "marital-status": "Married-
civ-spouse", "occupation": "Exec-managerial", "relationship": "Husband", "race": "White", "sex": "Male", "capital-gain":
 0, "capital-loss": 0, "hours-per-week": 13, "native-country": "United-States", "income": "<=50K", "Scored Labels": "<=5
0K", "Scored Probabilities": 0.26952520043708017}, {"age": 38, "workclass": "Private", "fnlwgt": 215646, "education": "H
S-grad", "education-num": 9, "marital-status": "Divorced", "occupation": "Handlers-cleaners", "relationship": "Not-in-fa
mily", "race": "White", "sex": "Male", "capital-gain": 0, "capital-loss": 0, "hours-per-week": 40, "native-country": "Un
ited-States", "income": "<=50K", "Scored Labels": "<=50K", "Scored Probabilities": 0.023948608105902526}, {"age": 53, "w
orkclass": "Private", "fnlwgt": 234721, "education": "11th", "education-num": 7, "marital-status": "Married-civ-spouse",
 "occupation": "Handlers-cleaners", "relationship": "Husband", "race": "Black", "sex": "Male", "capital-gain": 0, "capit
al-loss": 0, "hours-per-week": 40, "native-country": "United-States", "income": "<=50K", "Scored Labels": "<=50K", "Scor
ed Probabilities": 0.08759885505524728}, {"age": 28, "workclass": "Private", "fnlwgt": 338409, "education": "Bachelors",
 "education-num": 13, "marital-status": "Married-civ-spouse", "occupation": "Prof-specialty", "relationship": "Wife", "r
ace": "Black", "sex": "Female", "capital-gain": 0, "capital-loss": 0, "hours-per-week": 40, "native-country": "Cuba", "i
ncome": "<=50K", "Scored Labels": ">50K", "Scored Probabilities": 0.5060704221829483}]}}
THE END


C:\One\OneDrive - Lunar Cat\Z_Apresentacoes\20221112 - SQLSatBangladesh - MyFirstMLProject\Project\MyFirstMLProject\bin\
Debug\net6.0\MyFirstMLProject.exe (process 38208) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the conso
le when debugging stops.
Press any key to close this window . . .
```

Bad demo? Here's screenshots!



Questions?

"The unknown future rolls toward us.
I face it for the first time with a sense of hope, because **if a machine, a Terminator, can learn the value of human life, maybe we can, too.**"

Thank you!
धन्यवाद
Obrigado, pá!
Благодаря!
Дуже дякую!
धन्यवाद्
Tack så mycket!
¡Muchas gracias!
Vielen Danke!
Hvala vam!
ඔයාට ස්තූතියි
Ευχαριστώ!
Merci beaucoup!
Terima kasih!
Grazie mille!
Ďakujem!
Mulțumesc!
Labai ačiū!
Dziękuję Wam!
Mockrát děkuju!
Mange tak!
Kiitos!
Takk fyrir!
Dank u wel!
Takk!
Dank je!
Köszönöm!
Go raibh maith agaibh!
Diolch!

André Melancia
Andy.PT
LunarCat.PT