HACKWAGON
• ACADEMY •

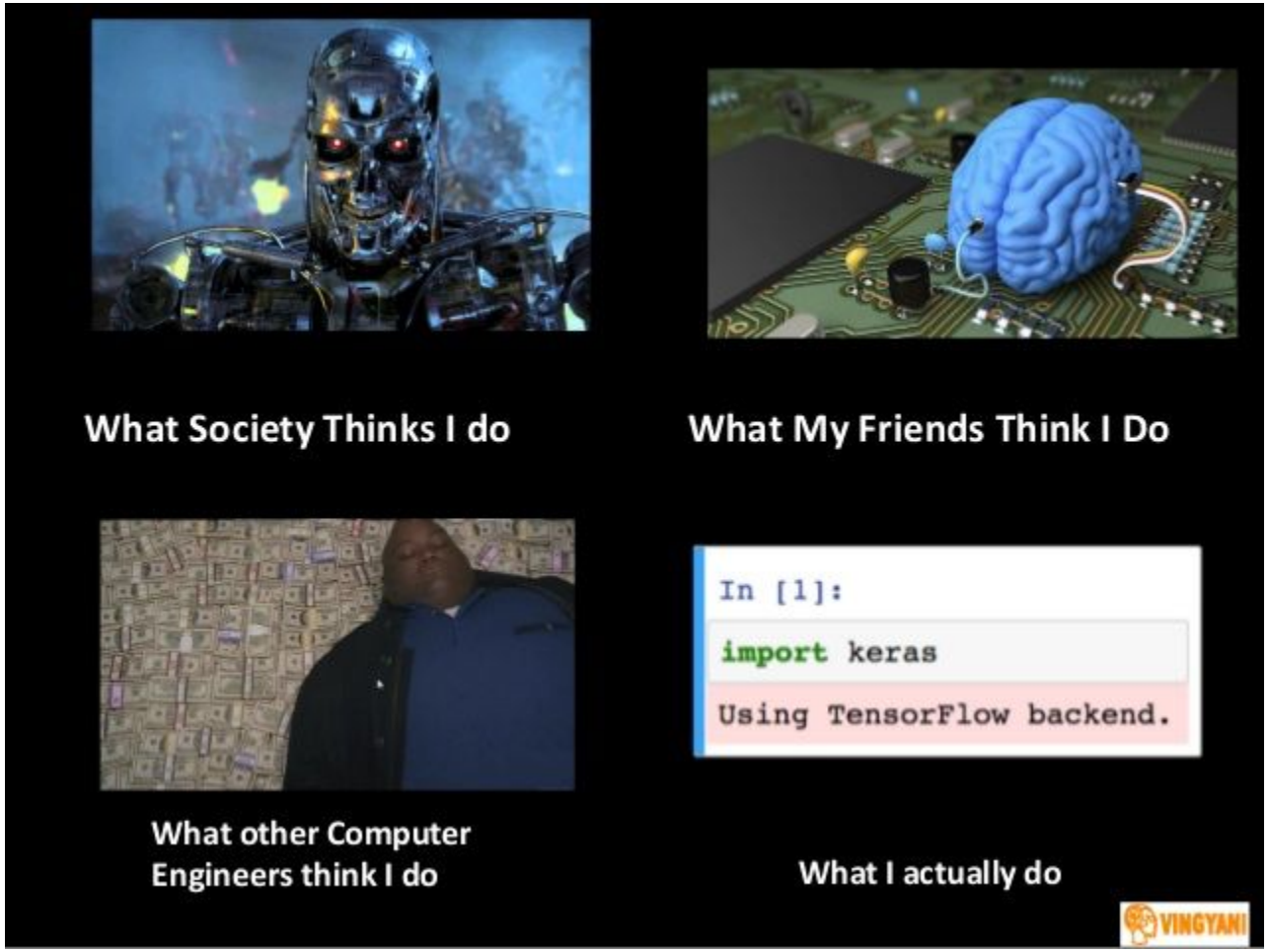
DATA SCIENCE 102:
INTRODUCTION TO MACHINE LEARNING

AGENDA

- What is Machine Learning?
- Types of Machine Learning
  - Applications of Machine Learning
- Sci-Kit Learn
  - Fundamentals
    - Standard steps
    - Reading the library
  - Examples
    - Regression

WHAT IS MACHINE LEARNING

WHAT IS MACHINE LEARNING



WHAT IS MACHINE LEARNING

A study of *algorithms and statistical models*

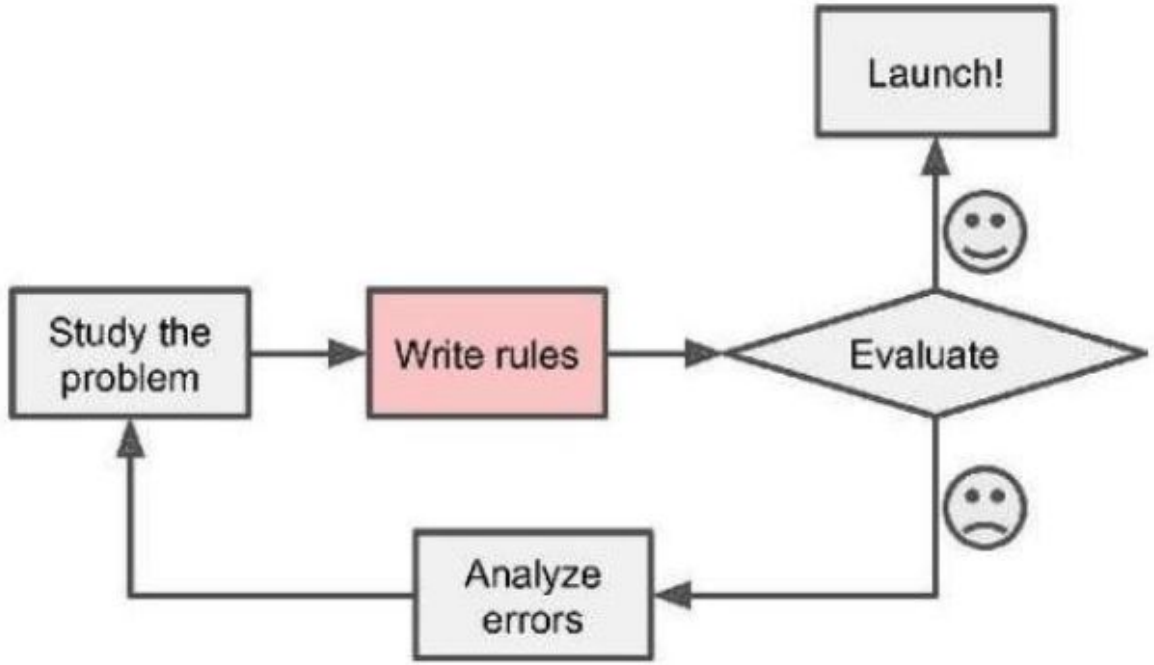
that computers use to perform a specific task effectively

without being explicitly programmed

EXAMPLE - SPAM EMAIL DETECTION

Without Machine Learning

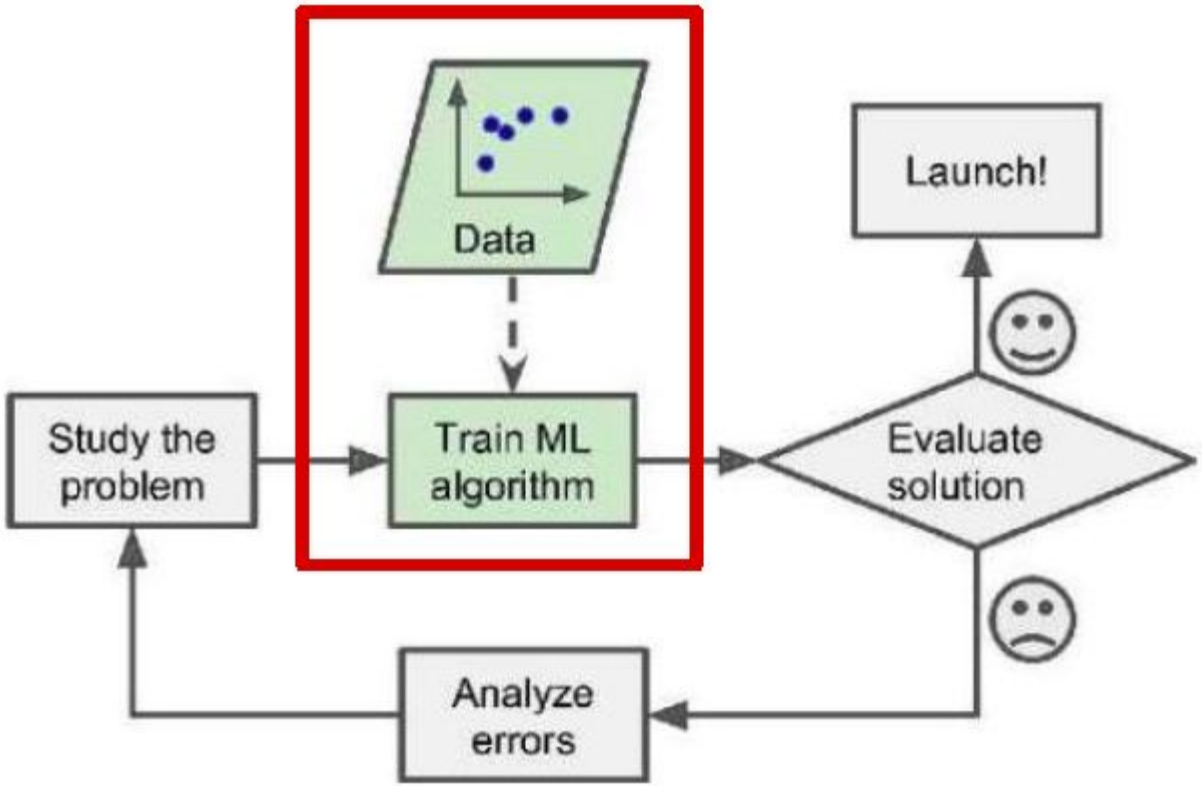
- You will write explicit rules to filter out emails containing words like:
  - *“4U”, “Loan”, “Prince”, “Money Transfer”, “Free”*
- Classify any email that contain these text as spam email



EXAMPLE - SPAM EMAIL DETECTION

With Machine Learning

- You train a model on some train data (spam & non-spam emails) with an algorithm that **infers** characteristics and patterns of spam emails
- The trained model can then apply learnt rules to detect signals of spam in an incoming email



TYPES OF MACHINE LEARNING

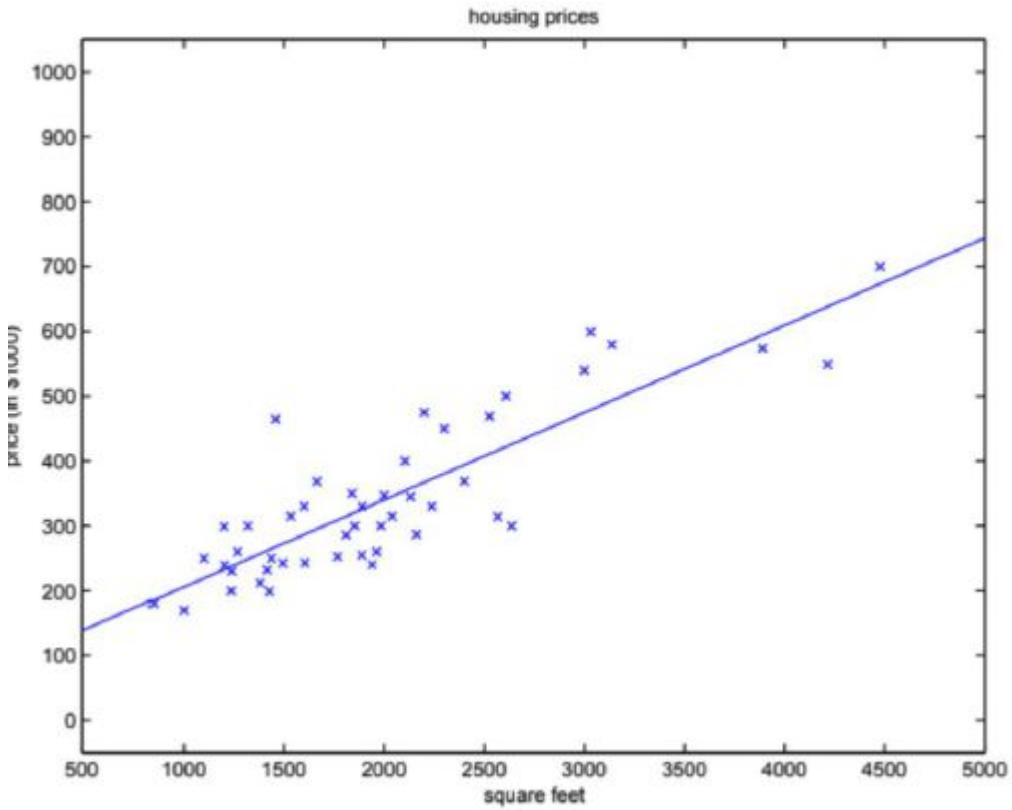
- Types of Machine Learning
- Supervised Applications
- Unsupervised Applications

TYPES OF MACHINE LEARNING

- Supervised Learning
  - Based on **labelled data,** makes **predictions** on a test set
- Unsupervised Learning
  - Based on **unlabelled data**, algorithm **discovers patterns** within the dataset

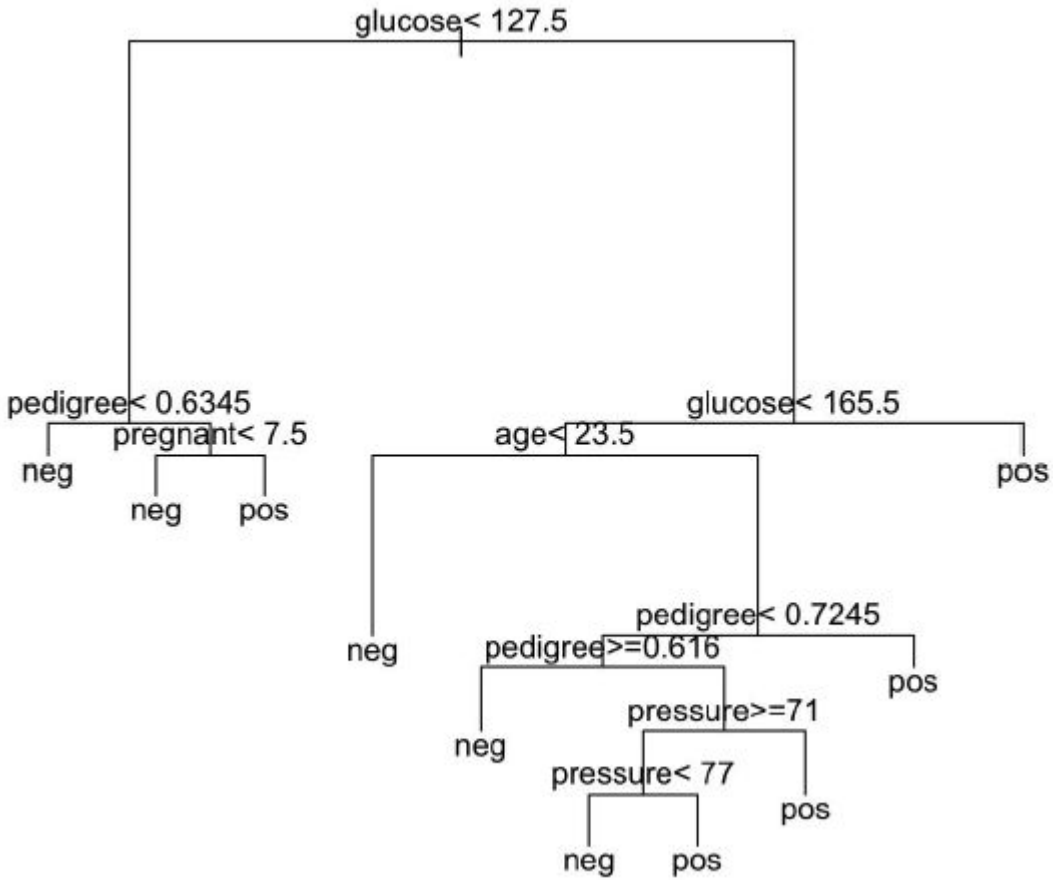
SUPERVISED APPLICATIONS

- Regression
  - Predict Housing Price (Target Variable) from Housing Floor Area (Input Feature)



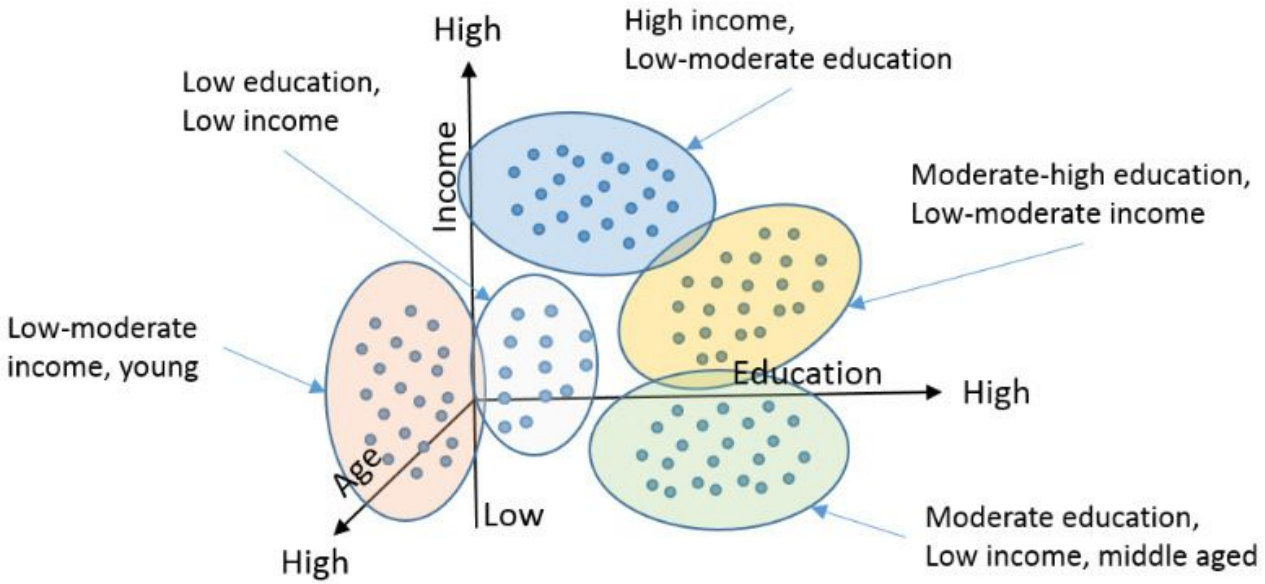
SUPERVISED APPLICATIONS

- Classification
  - Predict if a patient has diabetes (Yes/No)



UNSUPERVISED APPLICATIONS

- Clustering
  - Given a set of unlabelled data points, identify cluster each point belongs to



GENERAL MACHINE LEARNING STEPS USING SKLEARN

- Fundamentals
  - Standard Steps
  - Reading the library
- Examples
  - Regression
  - Clustering
  - Decision Tree

FUNDAMENTALS - STANDARD STEPS

When using any machine learning models with scikit learn, the following steps are usually applied in order:

Step 1: Choose a class of machine learning model from the library

Step 2: Choose the model's hyperparameters by instantiating with desired values (tuning)

Step 3: Arrange data into features and target

Step 4: Fit model to your data by using the `fit()` method of the model

Step 5: Apply the model to new data:

- For supervised learning, using the `predict()` method
- For unsupervised learning, using the `predict()` or `transform()` method

Python Data Science Handbook: Essential Tools for Working with Data by Jake VanderPlas (2016)

FUNDAMENTALS - READING THE LIBRARY

- Determine what model is used, then delve into details of that model in the library, and the other required parts of the library
- Here are some of the common classes you would use:
  - `sklearn.`<a name of a model> (i.e `sklearn.linear_model)`
    - Access to all functionalities of that model
  - `sklearn.metrics`
    - Functionalities of assessing the performance of your models
  - `sklearn.model_selection`
    - Performs cross validation and tuning of parameters
  - `sklearn.feature_selection`
    - Reduce dimension of dataset to boost performance

Python Data Science Handbook: Essential Tools for Working with Data by Jake VanderPlas (2016)

EXAMPLE - REGRESSION

```
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score

# Load the diabetes dataset
diabetes = datasets.load_diabetes()


# Use only one feature
diabetes_X = diabetes.data[:, np.newaxis, 2]

# Split the data into training/testing sets
diabetes_X_train = diabetes_X[:-20]
diabetes_X_test = diabetes_X[-20:]

# Split the targets into training/testing sets
diabetes_y_train = diabetes.target[:-20]
diabetes_y_test = diabetes.target[-20:]

# Create linear regression object
regr = linear_model.LinearRegression()

# Train the model using the training sets
regr.fit(diabetes_X_train, diabetes_y_train)

# Make predictions using the testing set
diabetes_y_pred = regr.predict(diabetes_X_test)
```

Step 3

Step 1 & 2

Step 4

Step 5