Day 1

Group 9

Price prediction for used cars

Kijiji -> Data mining -> their interface -> price of the car

Liniear regression

SQL DB

Clean data - remove noise - add 0

Group 15

Github issue label prediction

5 main types of issues

Text classification

clustering

Group 4

Violence in the News

Vategorize appropriate for kids

Vectorizers

Boyes trees revression, sum

Random forces

Group \_\_

NBA Expected Stats

Value of taking shot

Day2

Group \_\_

NHL Point progression

Age regression in prime?

pGPS developed to measure probability of success of prospects - developed by Jeremy Davis at Canucks Army

Clustering was done before it was known as clustering

Removed players with only a single season or not enough games (<15)

Group \_\_

Predictive policing (crime prevention)

Evaluated algorithm performance, gaussian NB, decision tree, Mulinomial NB, bernoulli np, perceptron, logistic regression, SVM

Group 13

Learning how to predict a MLB Most Valuable Player

Linear regression, ridge, lasso

Scraped MVP votind data from baseball-reference.com from 1961-2016

Group 14

Music “Hotttness” Prediction

Accurately predict the popularity of a song based on it’s characteristics

Used the million song dataset

Based it off of 12 different features

Removed unrated songs but kept songs with unknown features

Usd logistic regression, linear regression, support vector machines, gaussian Naive Bayes

For support vector machines: Support vector regression, RBD vs polynomial kernel, coefficient of deterination r^2=0.34

Classification algorithms have a higher success rate than regression