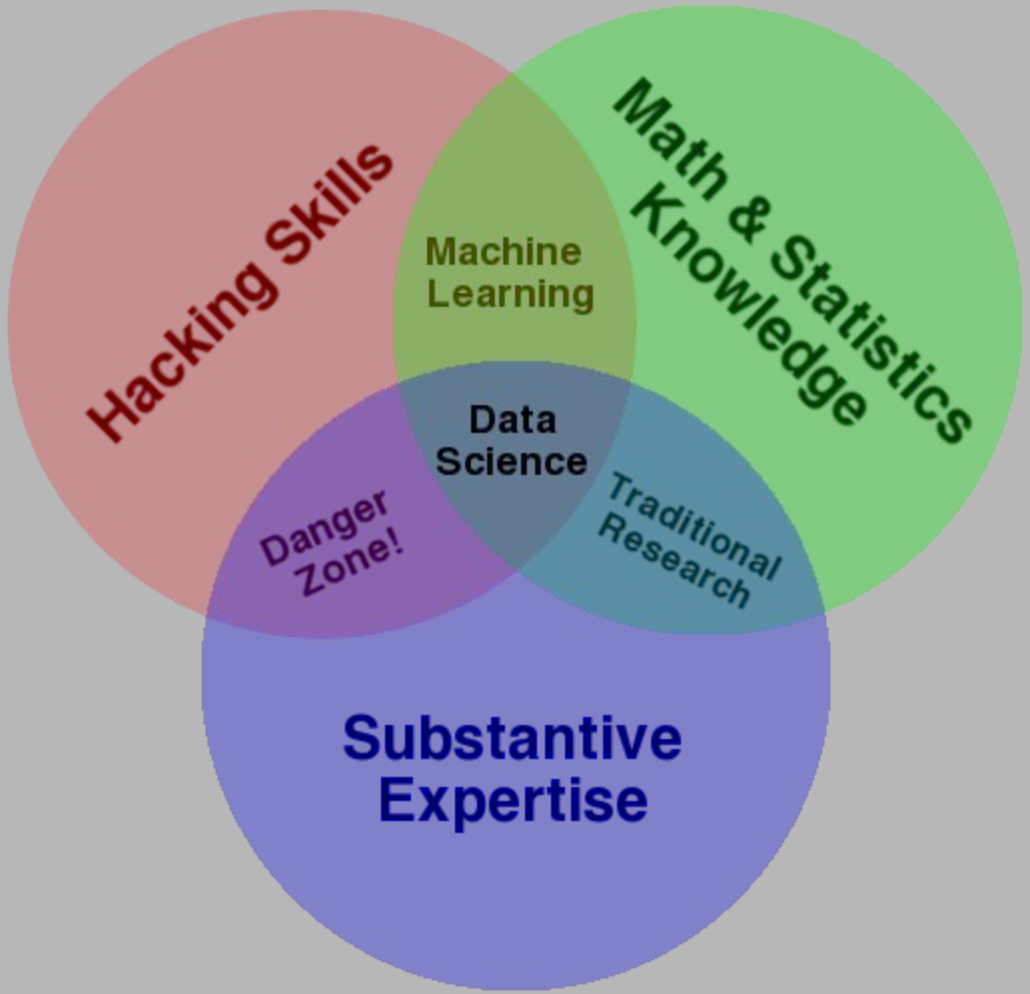
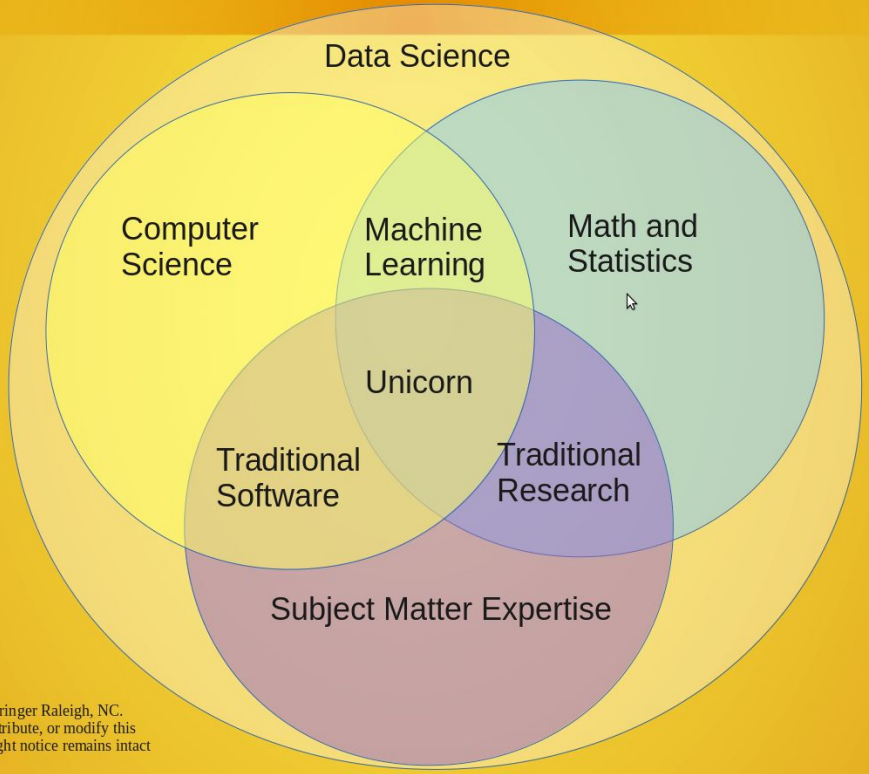
RunPassBot: A Data Science Project From Start to Finish

MEETING NOTES

- WORKING SESSION?
- UPCOMING MEETUP TOPICS
- LOCATIONS?
- SPEAKERS


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

Data Science Venn Diagram v2.0



WHAT IS RUNPASSBOT?

RUNPASSBOT IS A TWITTER BOT THAT:

- TRAINS A MACHINE LEARNING MODEL USING HISTORICAL NFL DATA
- TWEETS PROBABILITY PERCENTAGE OF A GIVEN PLAY BEING A RUN OR A PASS

EXAMPLE TWEET:

IND BALL @ GB 20:

RUN (67%)

WHY A TWITTER BOT?

EASY TO WRITE

COMPUTATIONALLY CHEAP TO RUN

API AND WRAPPER LIBRARIES

WHY NFL DATA?

OR

PICK A DATASET THAT YOU CARE ABOUT!

THE DATA SCIENCE PROCESS

1. FRAME THE PROBLEM
2. COLLECT RAW DATA
3. PROCESS DATA FOR ANALYSIS
4. EXPLORE DATASET
5. PERFORM IN-DEPTH ANALYSIS
6. COMMUNICATE RESULTS

1. FRAME THE PROBLEM

“CAN I PREDICT IF A TEAM WILL RUN OR PASS FROM THIS DATASET?”

2. COLLECT RAW DATA

INITIAL DATASET CAME FROM KAGGLE, THE REST CAME FROM USING NFLSCRAPR TOOL IN R.

3. & 4. PROCESS DATA FOR ANALYSIS /EXPLORE

THE DATA WAS SPREAD ACROSS SEVERAL CSV FILES

MANY FIELDS FROM THE DATASET WERE NOT NEEDED

NANs

5. PERFORM IN-DEPTH ANALYSIS

USED TPOT, A MACHINE LEARNING AUTOMATION TOOL TO SELECT MODEL FOR PREDICTION, IT SELECTED GRADIENT BOOSTING CLASSIFIER

SCORED 62% ACCURACY W/ 1 GENS

SCORED 64% W/ 5 GENS

6. COMMUNICATE RESULTS

PROBLEMS!

LIVE NFL DATA IS VERY EXPENSIVE! (CONTACT US RARELY MEANS CHEAP)

HALL OF FAME GAME WAS CANCELED