Predicting Counter Strike Global Offensive (CS:GO) E-Sport Tournament Match Winners

Matthew A. Lanham

PURDUE
KRANNERT
SCHOOL OF MANAGEMENT

About Experience

Data science and analytics is a hot topic and valuable skillset to have in today's data-driven world. E-Sports is on the rise with tournaments for CS:GO, League of Legends, Dota 2, and Starcraft creating new markets. Millions of fans from all over the world view watch these professional gamers play these games, and some are even gambling on the outcomes just like in other traditional sports. Ever wonder how to build a predictive model to predict whom is going to win? Professor Lanham is going to show you how to use Orange data mining software to do just that. Game on!

Download slides here:

About Me

Clinical Assistant Professor at Purdue Universities' Krannert School of Management and Co-Founder/Chief Data Scientist of Biz Analytics Lab, LLC in Lafayette, IN.

- Teach and mentor students (Fall/Spring semesters)
- Work with a couple fantastic partners (Summer)

At Krannert, course coordinator and teacher for:

- MGMT 571 Data Mining (Fall semester)
- MGMT 590 Using R for Analytics (Fall semester)
- MGMT 590 Predictive Analytics (Spring semester)
- MGMT 690 MS BAIM Industry Practicum (Spring semester)

Spend most of my time obtaining and mentoring experimental learning projects for students within Purdue's M.S. in Business Analytics & Information Management (BAIM) program.

Data Science, Business & Big Data Analytics

Big Data Analytics (BDA) (a.k.a. Data Science) is "*the practice of turning tools and raw data into something that non-data scientists might care about* (Ryza, Laserson, Owen, & Wills, 2015)."

Idea

- Analyzing big data provides a firm/business a path to extract new insights or create new forms of value which has potential to change markets, organizations, processes, etc. due to the large scale nature of the information (Mayer-Schönberger & Cukier, 2013).

“Small” versus Big Data Analytics

“Small” Data Analytics

- Where most companies are at today
- Using existing data stored in relational databases and data warehouses
- Structured datasets (numbers)

Big Data Analytics

- Where most companies want to get to
- Using new frameworks (e.g. Hadoop, Spark) to process large datasets faster
- Might contain unstructured datasets (words/text, images, video, audio)

Both should be focused on creating some sort of value

Business Analytics Breakdown

Core courses

MGMT 473 Data Mining

MGMT 306 Mgmt. Science

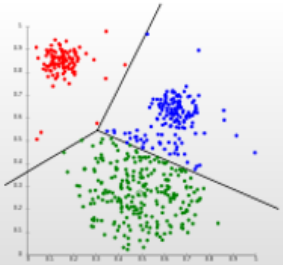
MGMT 474 Predictive Analytics

MGMT 472 Adv. Spreadsheet Modeling & Simulation

3 Types

Descriptive Analytics

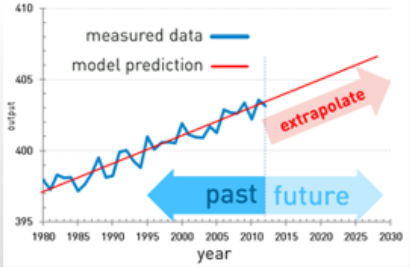
What has happened?



- Exploratory Data Analysis (EDA)
- Uni- and Multivariate Summaries
- Clustering - Segmentation - Profiling

Predictive Analytics

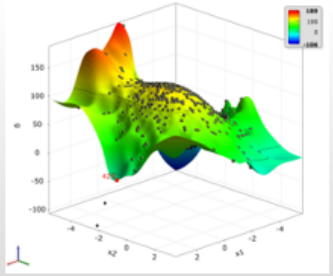
What will happen?



- Forecasting/Pattern recognition
- Classification
- Ensemble Modeling

Prescriptive Analytics

What is the best course of action?

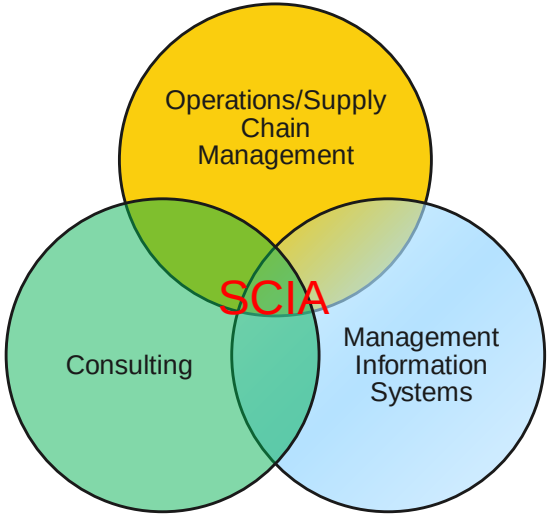


- Optimization/Heuristics
- Simulation
- Computational Stochastic Optimization

Business Analytics @ Krannert

- B.S. in **Supply Chain, Information and Analytics** (SCIA)
  - Intersection of Operations/Supply Chain Management, Analytical/Tech Consulting, and Management Information Systems (MIS)

- B.S. in **Industrial Management** (IM)
  - Hybrid program for mathematically confident students who aspire to mix their analytical skills and master all functional business areas (finance, marketing, etc.)



"A typical engineering student would have to finish their undergraduate program, work as an engineer for a few years and then go back to school for a business degree in order to get that technical management opportunity that I am pursuing. With the Industrial Management degree from Krannert, I will be able to essentially combine all those steps into one four-year undergraduate experience."
– Engineering to Krannert transfer student

Business Analytics @ Krannert

Supply Chain, Information and Analytics major

Analytical Consulting

$59,928 AVERAGE STARTING SALARY

SALARY RANGE **$40,000-75,000**

$7,667 AVERAGE BONUS (FOR THOSE RECEIVING A BONUS)

Management Information Systems

$59,200 AVERAGE STARTING SALARY

SALARY RANGE **$40,000-$101,000**

$4,056 AVERAGE BONUS (FOR THOSE RECEIVING A BONUS)

Operations and Supply Chain Management

$52,157 AVERAGE STARTING SALARY

SALARY RANGE **$32,000-$67,000**

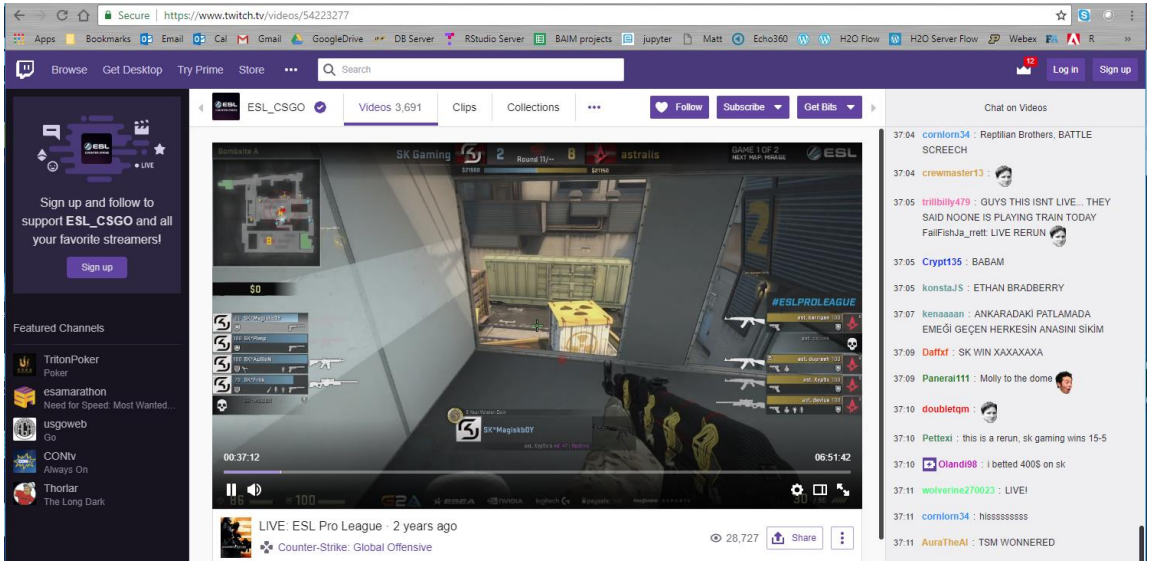
$8,357 AVERAGE BONUS (FOR THOSE RECEIVING A BONUS)

- Typical roles
  - Data Analysts
  - Managers
  - IT Consultants
  - Logistics Analysts
- Recent employers
  - Accenture
  - IBM
  - Deloitte
  - Microsoft
  - General Motors

Today's Analytics Challenge

Goal: Build and evaluate a predictive model to predict whom is going to win Counter-Strike: Global Offensive (CS:GO) professional matches

- Are we focused on Descriptive, Predictive, or Prescriptive Analytics?



Example: https://www.twitch.tv/videos/54223277

Quick survey:

- Who plays CS:GO?
- Who watches twitch.tv?
- Why do you watch?
- Who else cares about these matches and why?

Business Framing to Analytics Framing

First thing a good Analyst will do is understand the business problem. This entails answering at least these five questions:

1) What are the business goals/objectives?

2) Whom are the stakeholders and decision makers?

3) What decisions or actions can actually be taken?

4) What is the success criteria or key performance measures (KPIs)?

5) What constraints or restrictions exist?

Business Framing to Analytics Framing

Let me give you some context…

- There was an e-sports data hackathon sponsored by Sportradar, Scout, and Loci
- Goal was to use any of e-sports Sportradar data to develop something potentially novel

Lets go back to our questions (you need not answer these in sequential order):

1) What are the business goals/objectives?

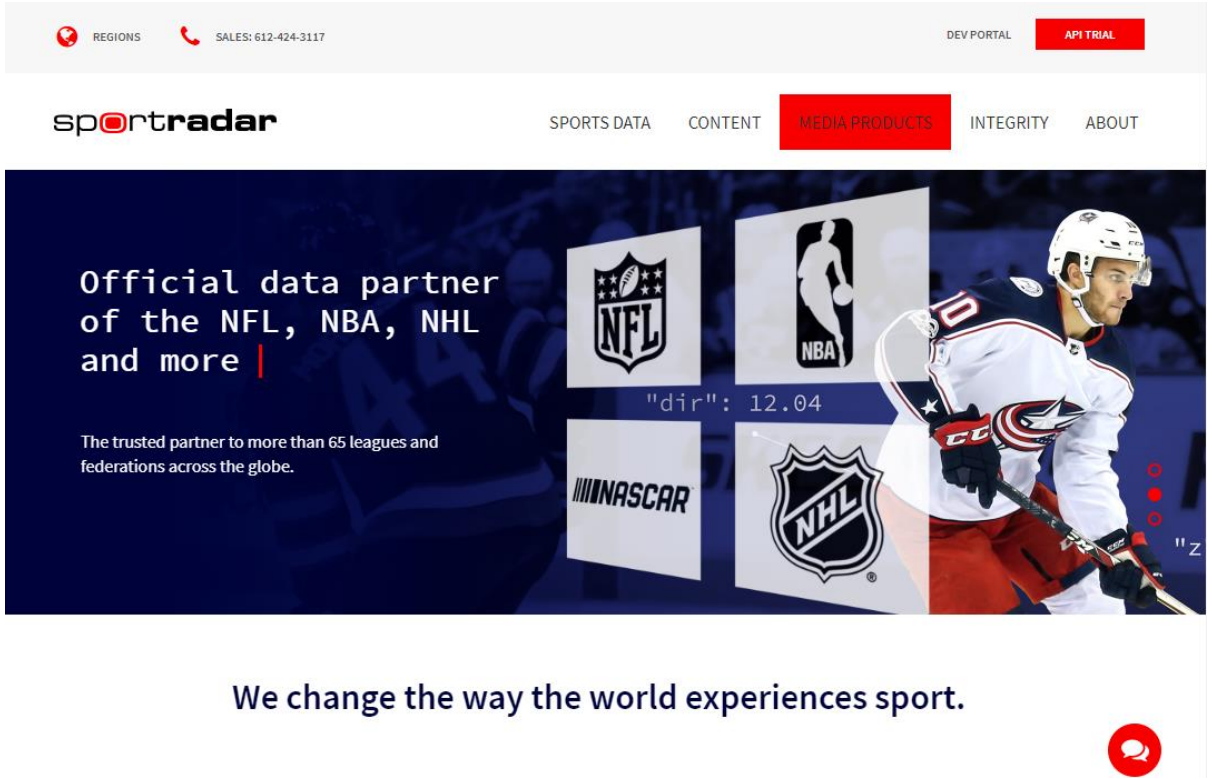
2) Whom are the stakeholders and decision makers?

3) What decisions or actions can actually be taken?

4) What is the success criteria or KPIs?

5) What constraints or restrictions exist?

Who cares?

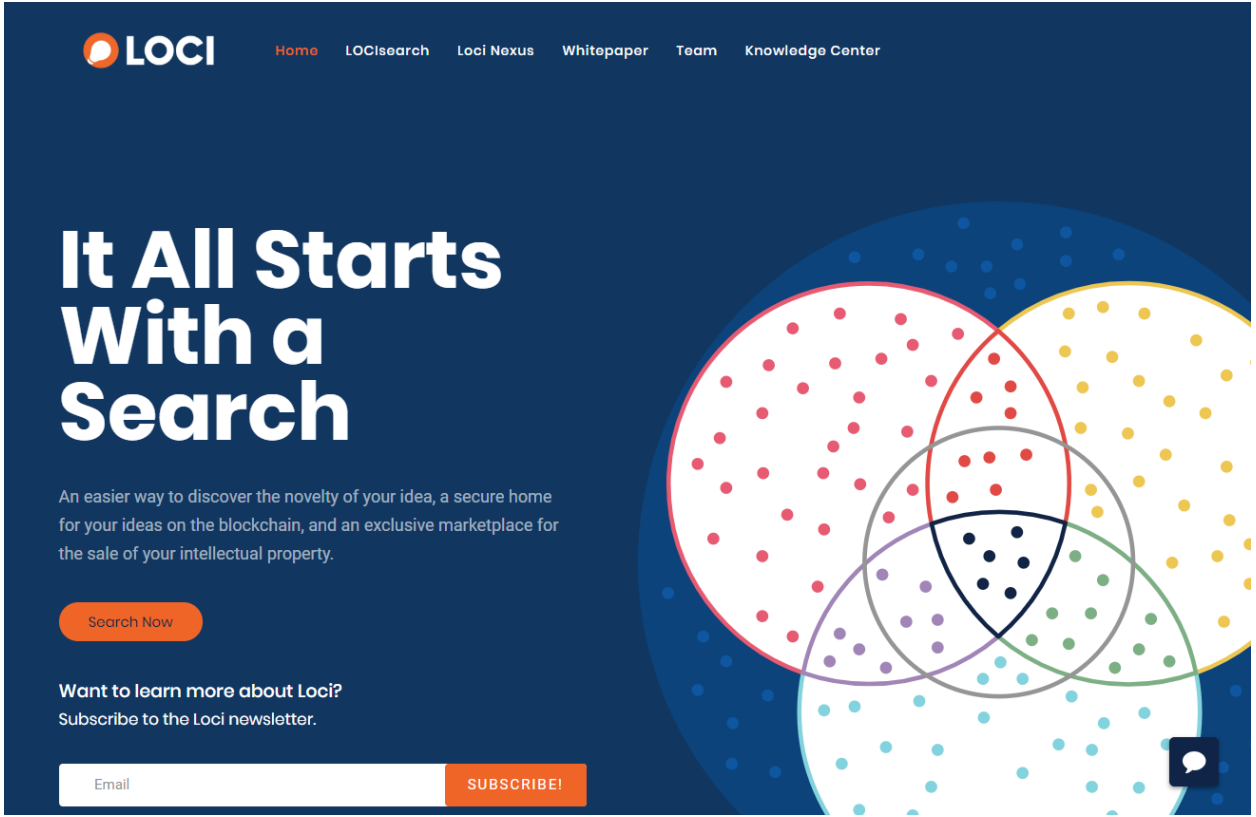


https://sportradar.us/

Who cares?

http://scoutesports.com/

Who cares?



https://loci.io/

Framing the Problem – eSports History

The eSports boom began in the second half of the 2000's:

Full acceleration on eSports becoming a mainstream form of media started in the 2010's:

Both these games were subject of significant tournaments and grew viewership substantially (e.g. LoL has over 100 Million players today)

Framing the Problem – eSports Market

As of 2017 there are over **130 million** electronic sports (eSports) fans and enthusiasts (Lee & Schoenstedt).

Source: SCCG Management (2017)

- eSports market is currently valued at approximately **$900 million** (Business Insider, 2018)
- The online gambling market is around $56B.
- eSports betting is approximately **$6B (11% of online)** (Grove, 2016).

eSports Betting

Betting on eSports began with *Counter-Strike: Global Offensive (CS:GO)* through a process called "skin betting" or "skin gambling."

The skins themselves provide no advantage, but gamers love to customize their experience and show off to others.

eSports Betting

Different types of bets can be placed in the eSports betting environment, and these include

- Outright Bets
- Match Bets
- Handicap Bets
- Accumulator Bets
- Special Bets

eSports Betting

Outright Bets

Case 1: Bet placed on a player or team in tournament.

Case 2: Bets placed on statistics rather than overall game outcome (e.g. most kills within the tournament)

eSports Betting

Match bets (*most common bet*)

Bets placed on a single one-on-one match.

Modes: Depends on bookkeeper

1) Cash
2) Skins (*CS:GO is most popular game for this*)
3) Cryptocurrency

3 options: Win, Lose, or Draw.

eSports Betting

Handicap Bets

- places a minus handicap against the player that is the favorite to win, whereas player who is less likely to win gains a positive handicap.
- Idea: player less likely to win gets a head start, and the favorite must play catch-up.
- Typically, odds for both players will be around 1.9.

eSports Betting

Accumulator Bets (i.e. the accumulation of bet)

- Rather than waiting for one wager, you are waiting for multiple simultaneously.
- An accumulator is riskier, but those risks yield a greater reward.
- This is often where large sums of money are won. It's not rare to see people winning thousands from bets of $2- 3.

eSports Betting

Special Bets/Novelty Bets (least common for eSports)

- They offer odds on something that is very unlikely to happen within a tournament.
- Often bets placed just for a bit of harmless fun to make an event more interesting and with decent odds. (e-sports betting, 2017)

eSports Betting Market

Source: SCCG Management (2017)

This year, eSports betting is expected to **surpass $6B** (Grove, 2016).

Bets dropped down to from 5.5B to $3B due to the restrictions **valve** placed on skin betting.

eSports Betting Market

HOME | VIRGINIA | NATIONAL | ISSUE AREAS | 2018 RACES | FEATURED A

Home › Sponsored Content › Law Banning Sports Gambling Defeated by Supreme Court

SPONSORED CONTENT SPORTS

Law Banning Sports Gambling Defeated by Supreme Court

By **Blue Virginia** - July 22, 2018 153 0

SPONSORED CONTENT

Sports fans across the country got some great news in May when the United States Supreme Court struck down the law that bans sports gambling with a 6-3 ruling. Now, states can take the proper measures to legalize sports betting as they see fit. The law that was defeated, known as the Professional and Amateur Sports Protection Act, was originally enacted in 1992.

The law only allowed individual game betting in the state of Nevada. In fact, sportsbooks in the state saw a record $4.8 billion in wagers placed in 2017 alone. The states that plan to move quickly to legalize sports betting are Delaware, Pennsylvania, New Jersey, West Virginia, New York and Mississippi.

Article link

Expect a wave of changes to the law regarding gambling very soon…

GAMBLING IN THE USA

Caesars' entry boosts sports betting in Indiana

Published 3 weeks ago on July 4, 2018
By Niji Ng

Article link

Business Problem – eSports Betting Market

92% of all current bets in cash gambling come from these 4 games:

38% market share

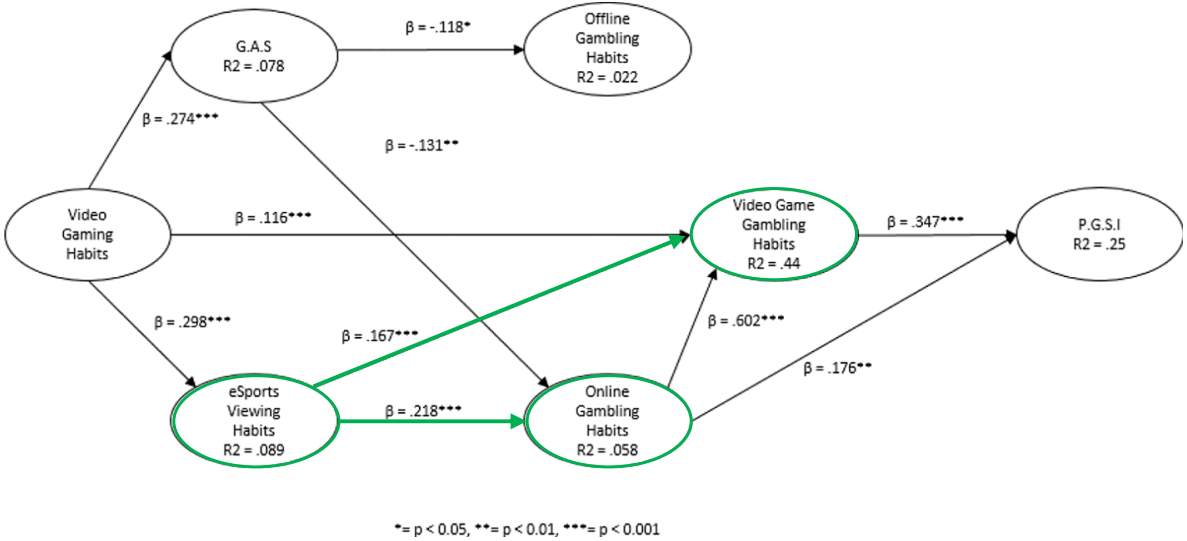
29% market share

18% market share

7% market share

Relationship among gamers, viewers, and betters

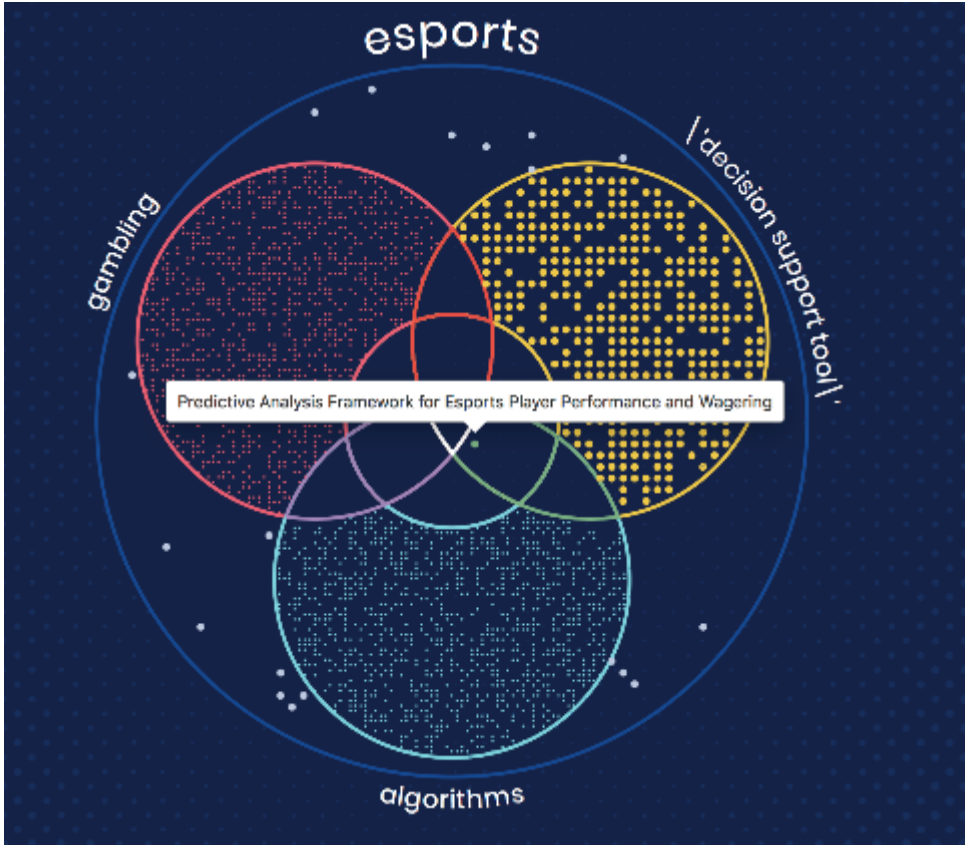
Macey & Hamari (2018) "Investigating relationships between video gaming, spectating esports, and gambling"



- *Video Gaming Habits* are a moderately strong predictor of esports consumption (β = 0.298, p = <0.001).
- Moderate effect *on Esports Viewing Habits* (β = 0.218) and *Online Gambling Habits* (β = 0.167) on *Video Game-Related Gambling Habits*

Potential for this prototype to be patented?

According to Loci….



Datathon Workflow

a) CS:GO

~~b) Odds Comparison~~

c) Tournament Summaries

Data Exploration/ Descriptive Analytics

What has happened in the past?

Predictive model(s)

What will happen in the future?

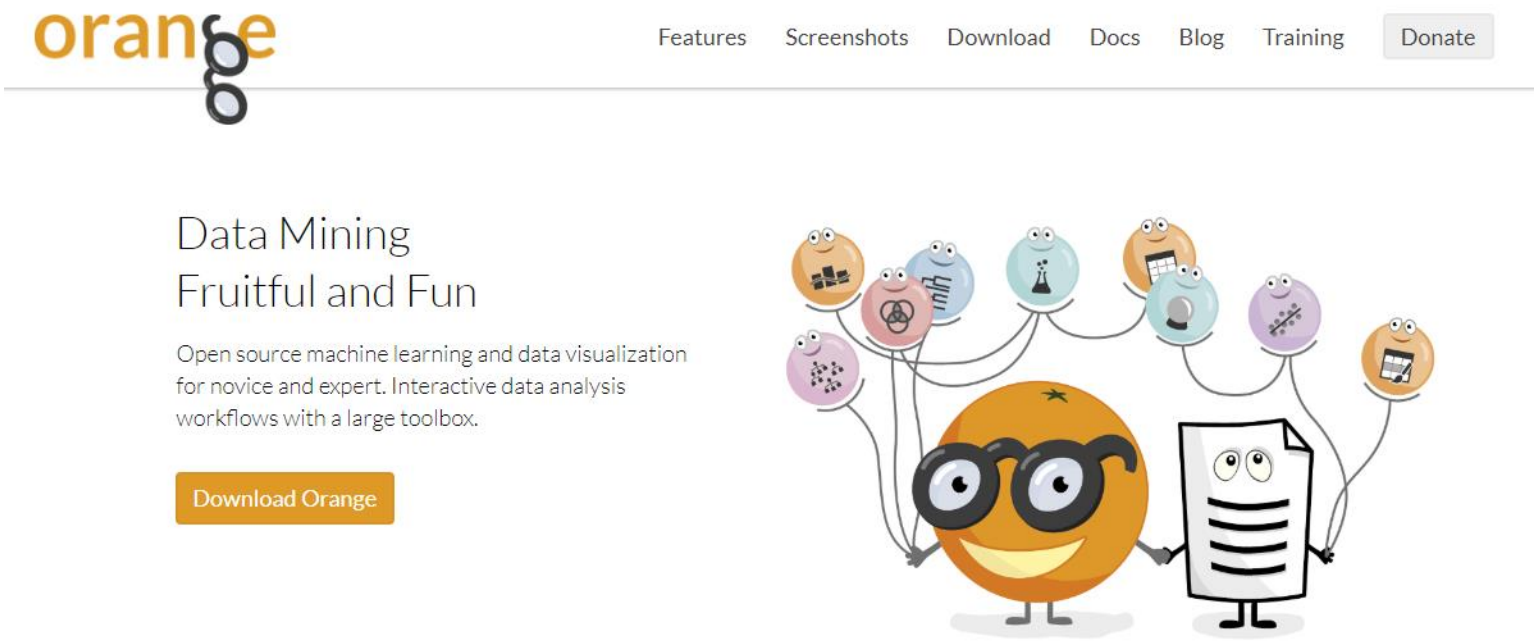
We'll use Orange

Prescriptive model

What can you do about it?

Orange software

https://orange.biolab.si/



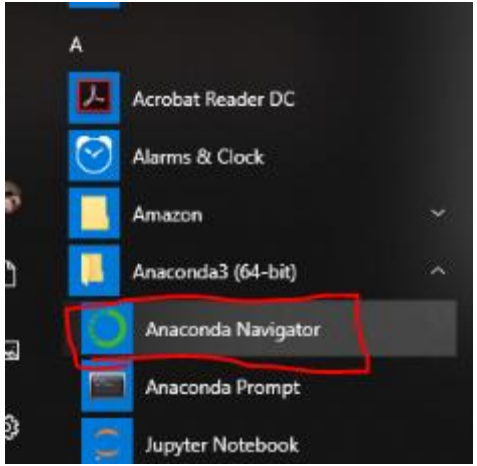
Orange is a nice tool to learn and explore, and has a feel similar to software we actually teach/use in business (e.g. SAS Enterprise Miner, IBM SPSS Modeler, Microsoft Azure)

Orange software – YouTube example videos

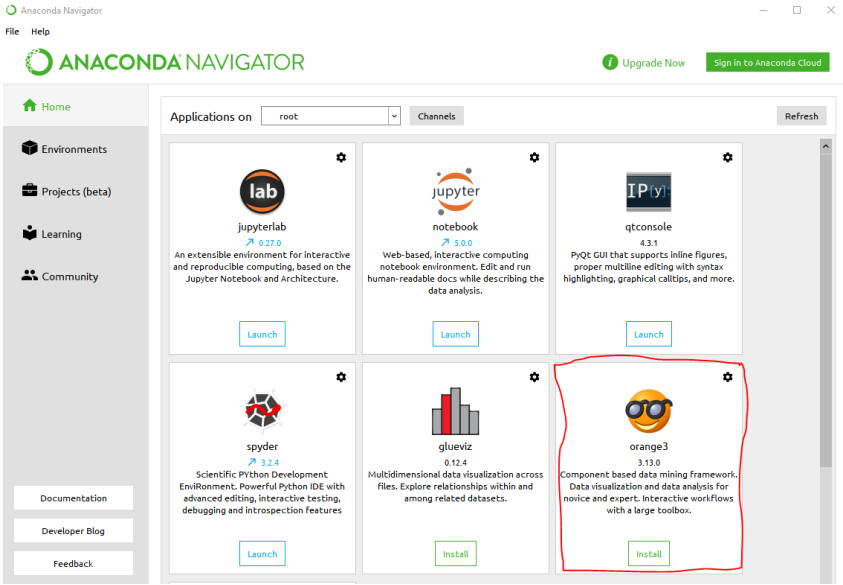
https://www.youtube.com/watch?v=HXjnDIgGDuI&list=PLmNPvQr9Tf-ZSDLwOzxpvY-HrE0yv-8Fy

Open Orange Software

1) Go to Start Menu
2) Find Anaconda 3 (32 or 64 bit – does not matter)
3) Select Anaconda Navigator

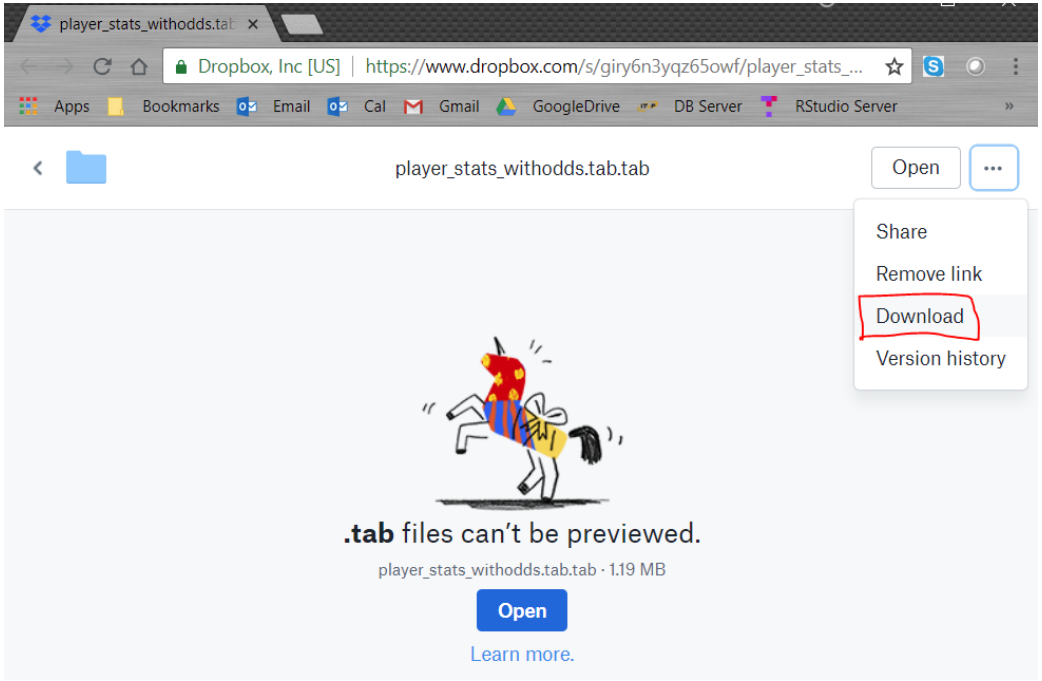


4) Launch Orange 3



Download CS:GO dataset to your desktop

https://www.dropbox.com/s/giry6n3yqz65owf/player_stats_withodds.tab.tab?dl=0



Data Exploration

Tournaments

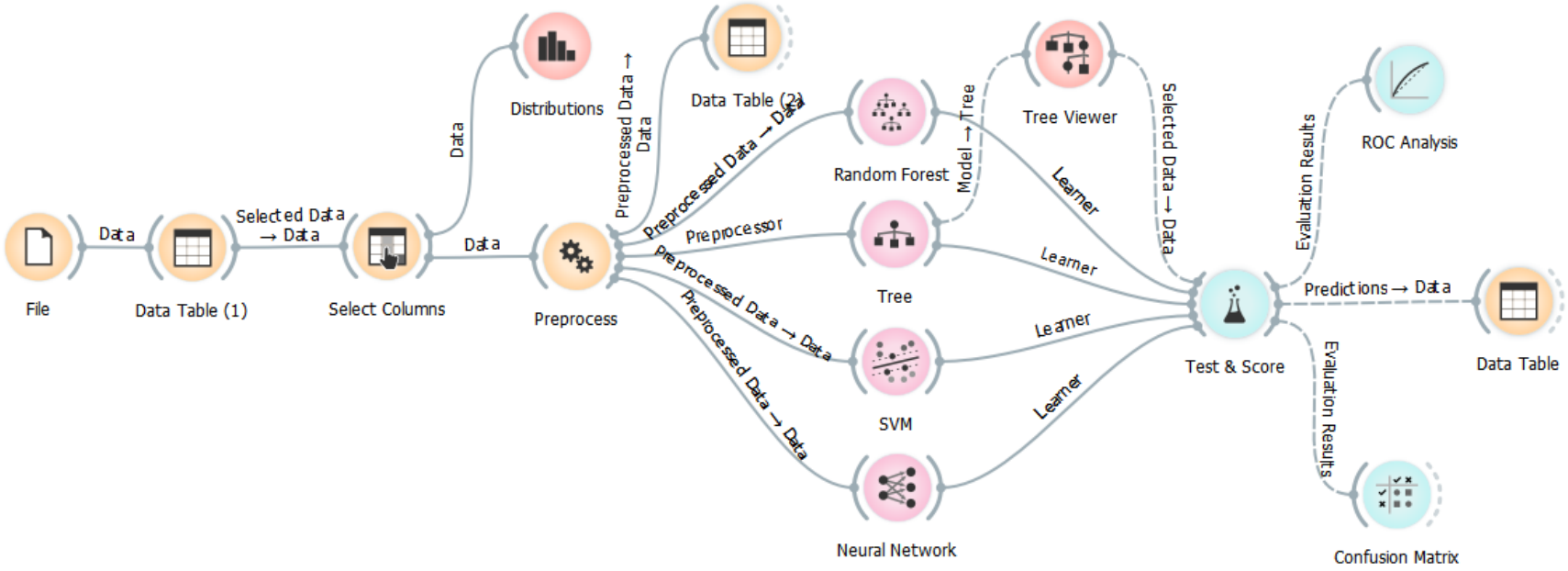
Tn Name	Scheduled Time 2017 Q1	2017 Q2	2018 Q1	2018 Q2
ESL Pro League			26	38
Intel Extreme Masters			10	16
ESL ANZ Championship			16	14
ESL Meisterschaft			10	4
ESL UK Premiership			10	3
ESL Benelux Championship				8
ESL Championnat National				8
ESL Mistrzostwa Polski			7	8
ESEA Premier Division		8		
ESL Brasil Premier League	4	8		

- 10 different tournaments
- Has taken off from 2017 to 2018

The data gives more support that these tournaments will continue to grow..

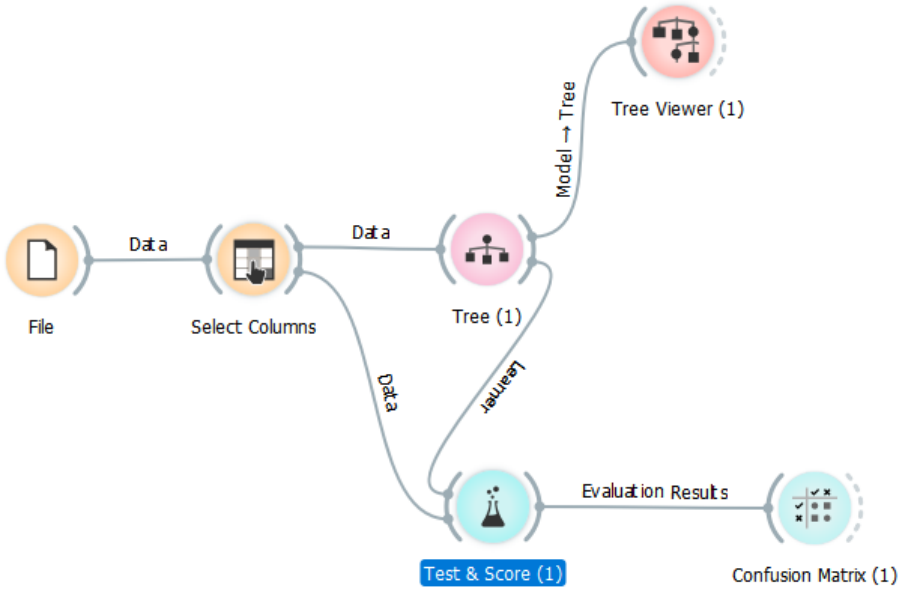
Explore our Orange Flow

This Example Orange Flow allows use to visually see step-by-step what we did in our analysis. In this flow there are multiple models being built and compared.



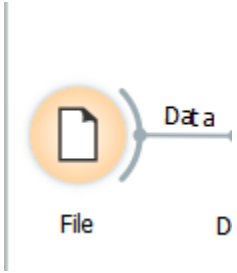
Explore our Orange Flow

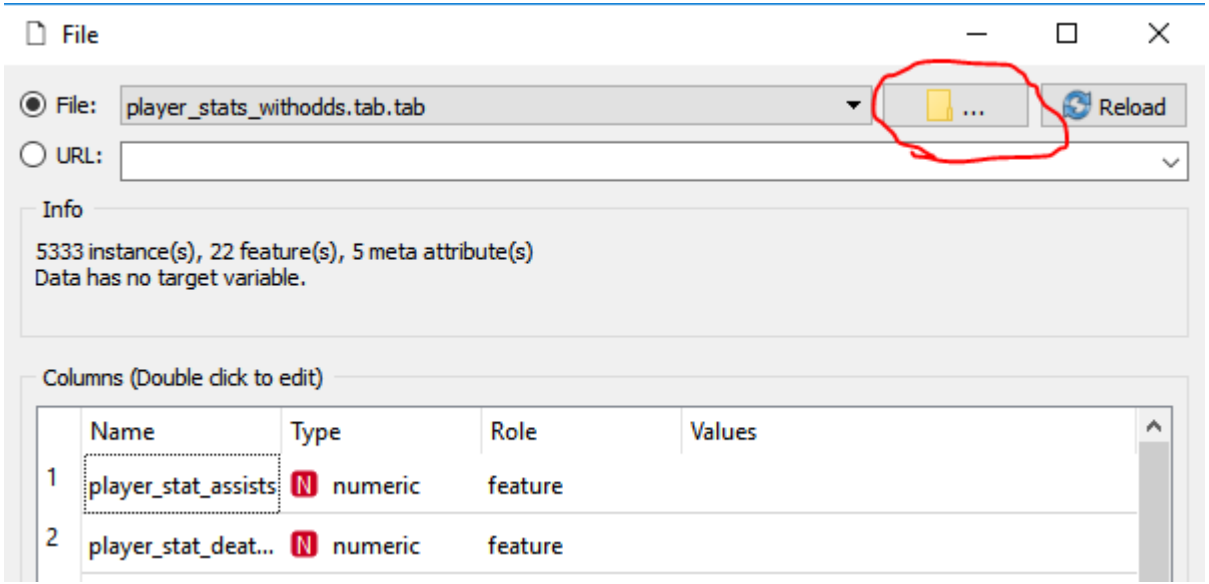
We'll use this basic classification tree model to predict team winners.



Load in dataset

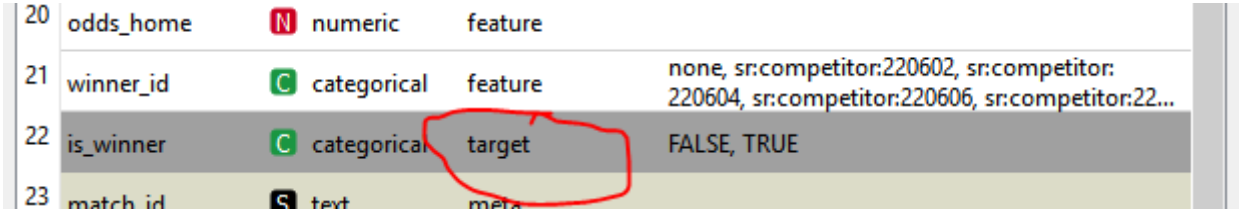
Right click on File





Target variable role

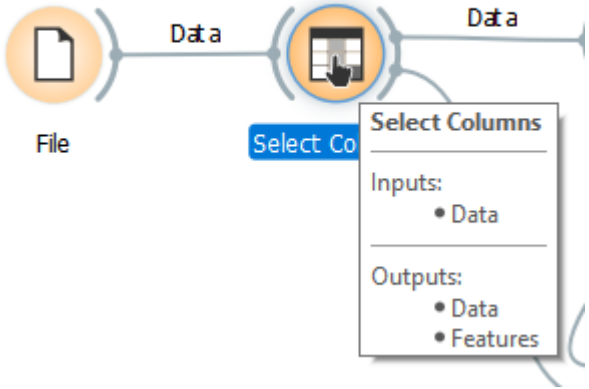
Make sure is_winner is the “target” variable. This is what we are trying to predict.

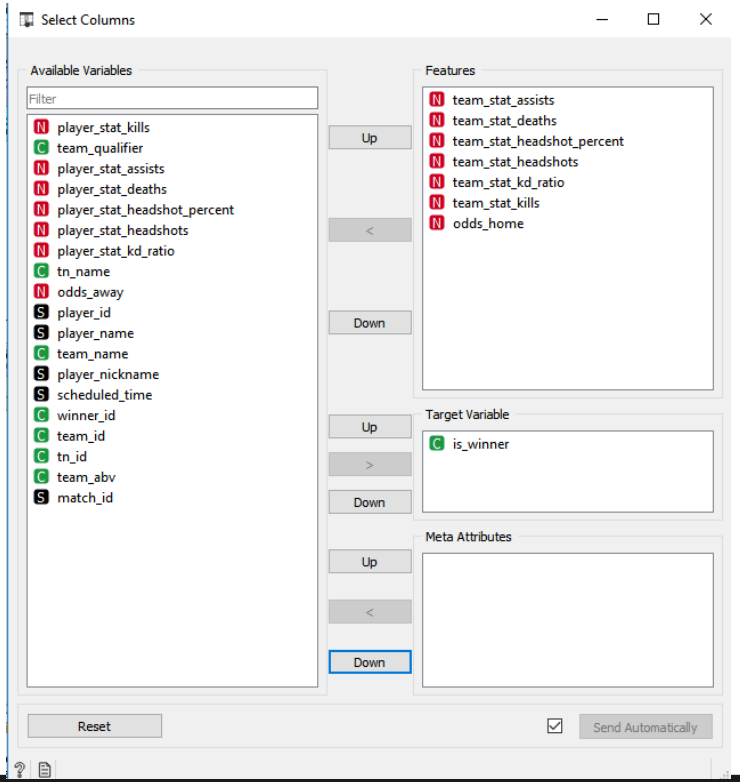


20	odds_home	N numeric	feature	
21	winner_id	C categorical	feature	none, sr:competitor:220602, sr:competitor:220604, sr:competitor:220606, sr:competitor:22...
22	is_winner	C categorical	target	FALSE, TRUE
23	match_id	S text	meta	

Target variable role

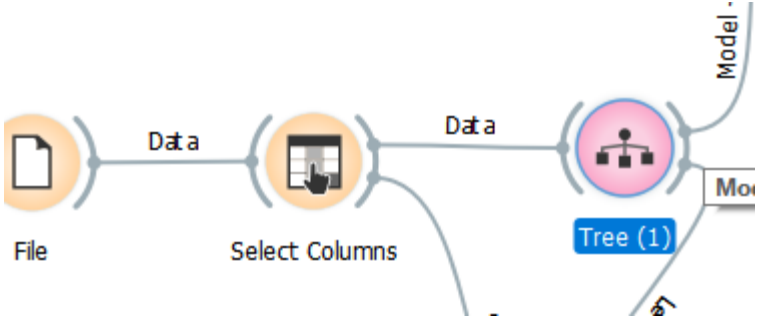
Right click on Select Columns and lets tell the software which inputs we want to use to predict team winners





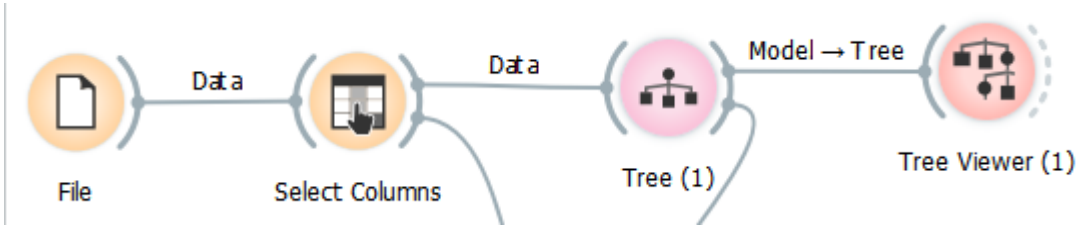
Now lets feed this into a predictive model

If you right click on the Tree, there are settings you can change that might impact the performance (how well the model is able to predict). We'll just use the default settings.



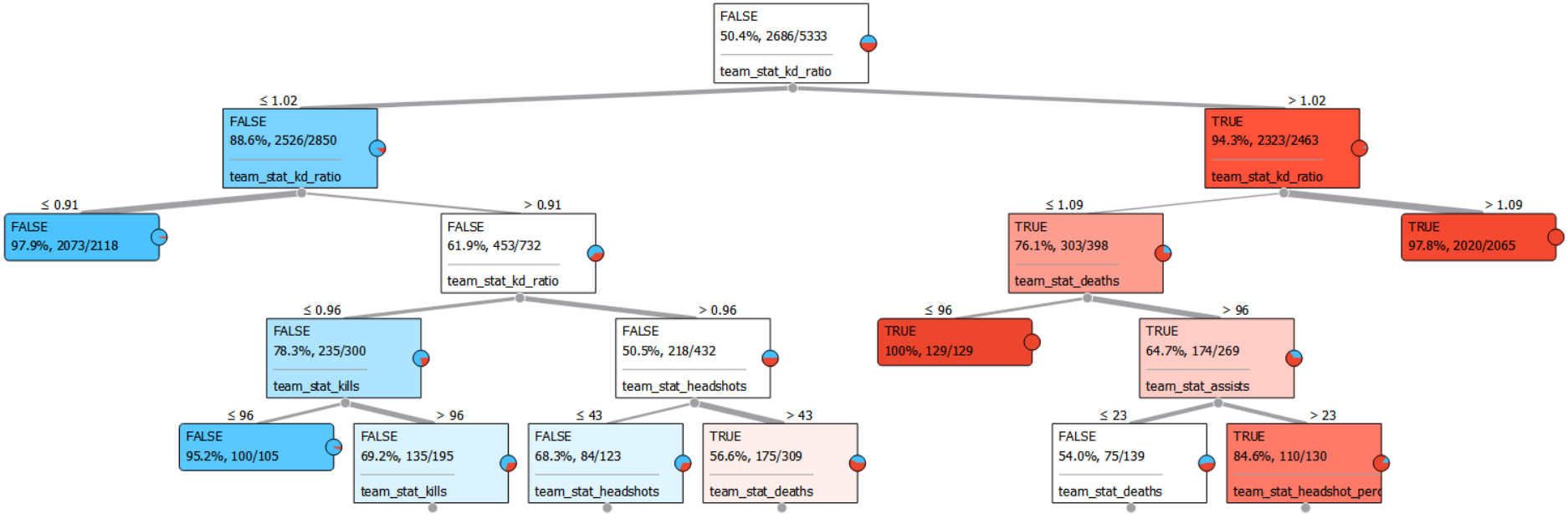
We can visualize a decision tree model

Double-click the Tree Viewer and you can see what features/inputs were used to predict team winner.



Decision tree model

Which features are important?

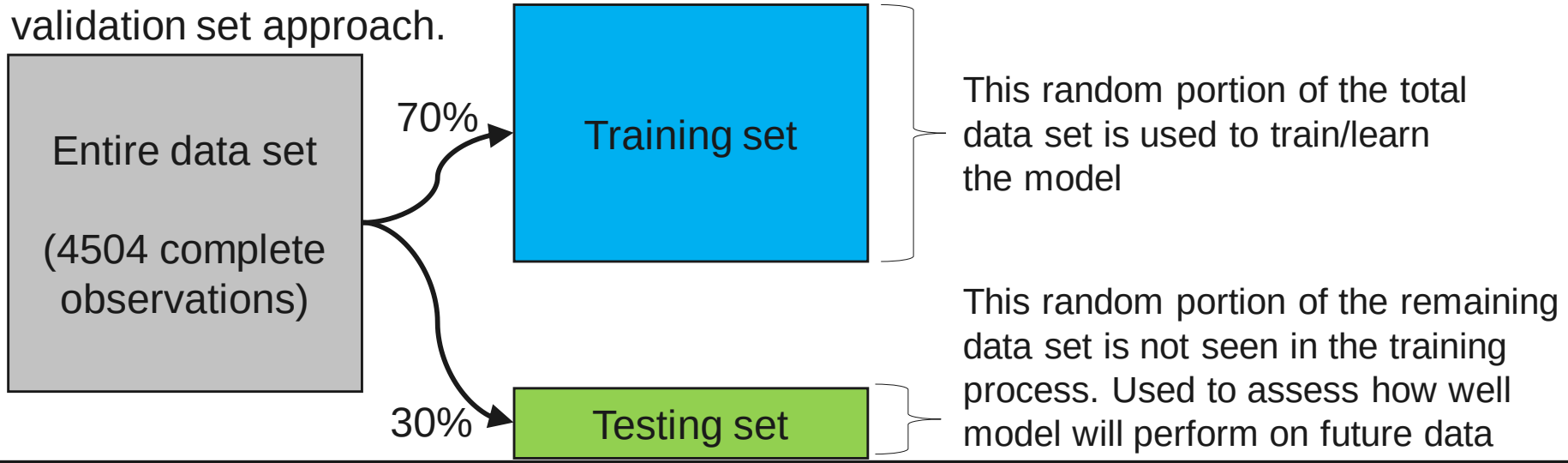


Predictive Model Design – Cross validation

The idea of **cross-validation** is establish a more realistic measure of accuracy (or error) of how your model will actually perform on future observations.

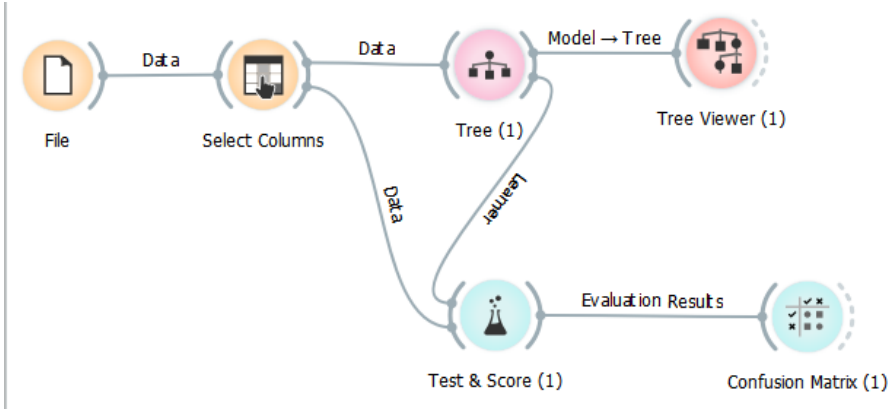
- You don't want to use the entire data set to both train/learn your model, as well as evaluate how well it performed.

There are various approaches to perform cross-validation. This is the traditional validation set approach.



Model evaluation

Use the Test & Score widget to see how well our model does on future observations (i.e. the test data set).



Test & Score (1)

Sampling

Cross validation

Number of folds: 2

Stratified

Evaluation Results

Method	AUC	CA	F1	Precision	Recall
Tree	0.975	0.953	0.953	0.953	0.953

This table shows various statistical performance measures that held us evaluate how well our model performs.

Model evaluation – confusion matrix

The confusion matrix is a 2x2 table that counts how many times your model classified winners versus losers correctly.

		Predicted		
		FALSE	TRUE	Σ
Actual	FALSE	2558	128	2686
	TRUE	125	2522	2647
	Σ	2683	2650	5333

		Predicted		
		FALSE	TRUE	Σ
Actual	FALSE	95.3 %	4.8 %	2686
	TRUE	4.7 %	95.2 %	2647
	Σ	2683	2650	5333

Well, what do you think? Is this model good enough to gamble with?

Remember these questions??

Lets discuss these a bit, so you can get a better feel for how you'll learn to think as a business analyst...

1) What are the business goals/objectives?

2) Whom are the stakeholders and decision makers?

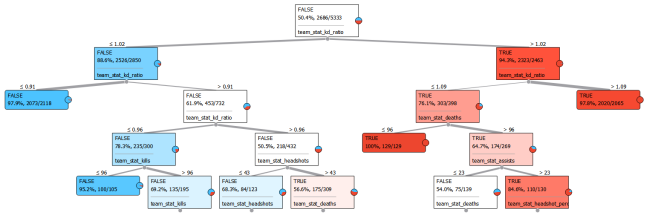
3) What decisions or actions can actually be taken?

4) What is the success criteria or key performance measures (KPIs)?

5) What constraints or restrictions exist?

Conclusions

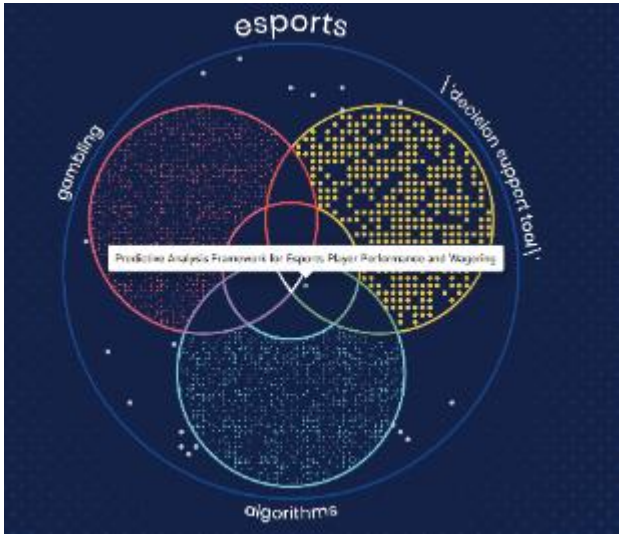
Value proposition 1: Scout tries to create community value and new careers for people into eSports. The current focus is on helping gamers improve and identify talent. Having the ability to predict match performance is one step to identify causes of good/poor performance.



Value proposition 2: An eSports better might want to know how to have the ability to identify potential arbitrage opportunities in the gambling market and/or have a decision support tool to help them bet/invest more intelligently. We believe this would increase your customer base and involvement.

Conclusions

Value proposition 3: We showed that using Loci, such a betting decision support system for e-Sports is indeed novel and could be patented.



Undergraduate Analytics Projects

2018 Purdue Undergraduate Research Conference (April 2018)

- Check out their project posters at MatthewALanham.com
- We won 1st and 3rd place
- To work with me, you just need to have taken Data Mining or Predictive Analytics first

Questions About Analytics @Krannert?

Bookmark my website: MatthewALanham.com

Follow me on Twitter: @MatthewALanham

Email me: lanhamm@purdue.edu