WIZELINE®

AI IN ACTION:

Applying Machine Learning Models to the 2018 World Cup™

© 2018 All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise, without full attribution.

AI IN ACTION:

Applying Machine Learning Models to the 2018 World Cup™

The 2018 FIFA World Cup™ took the summer by storm. Soccer fans around the world tuned in to watch the highly anticipated tournament in Russia. This case study dives into Wizeline's experiment to predict the winner of the 2018 FIFA™ tournament using predictive algorithms. The result? A tool that outperformed 95% of human participants. We'll share the conditional probabilities used by Wizeline and review the Machine Learning algorithms that companies can apply to solve unique business problems.

In just a short period of time, machines have proven to be more savvy at predictive modeling than their human counterparts. While some people fancy themselves visionaries, capable of outsmarting even the best odds, the majority rely on well-vetted data from experts before placing their bets.

We set out to build a predictive algorithm that could compete with the best of them. The 2018 FIFA World Cup™ presented the perfect opportunity to put it to the test. It also presented an opportunity to foster friendly competition and team culture at Wizeline.

Our data scientists eagerly accepted the challenge of creating a prediction tool to estimate the probability of each country advancing, and ultimately winning the World Cup. We named it Paul, after the famous octopus that "predicted" the outcome of the 2010 FIFA World Cup™ in South Africa. Our Paul, however, has neatly defined variables, scientific hypotheses, and operates based on data, not chance.

First things first: What is AI?

Artificial Intelligence (AI) is the concept that machines are able to carry out tasks in a smart way. **Machine Learning (ML)** is an application of AI, and the fastest growing area of AI. The rapid rise of AI technologies over the last ten years has been largely due to advances in Machine Learning. It consists of building algorithms that learn from experience and make predictions about data. **The main idea behind Machine Learning is that machines can learn for themselves if we give them access to the data.**

In this case, AI is a program that mimics human characteristics, like predicting a match score. It takes the form of computerized decision-making algorithms.

Setting the Framework

Our Model

First, we looked at the goals scored and the match outcomes of the countries participating in the World Cup™ over a two-year period. Our model only uses data from non-friendly matches, because we believe these are a better representation of a team's actual prowess.

Goal Intensities and Match Outcomes

The first step to predicting the outcome of a match is to estimate the expected number of goals team A scores against team B. Unfortunately, many of the countries facing each other in the World Cup™ have not played each other in recent time. There are relatively few matches between national teams in football, and hardly any between countries of different confederations. To circumvent this, we looked at the outcome of recent matches of team A, focusing on the number of goals they scored. This data is then triangulated by factoring in the relative defense of team B with respect to each of their opponents.

Friendly match – an exhibition game that has no impact on a player or team's ranking, or in which the impact is greatly reduced. Commonly referred to as a scrimmage or preseason game.

Non-friendly match – an official match which directly impacts a player or team's ranking in a league or tournament.

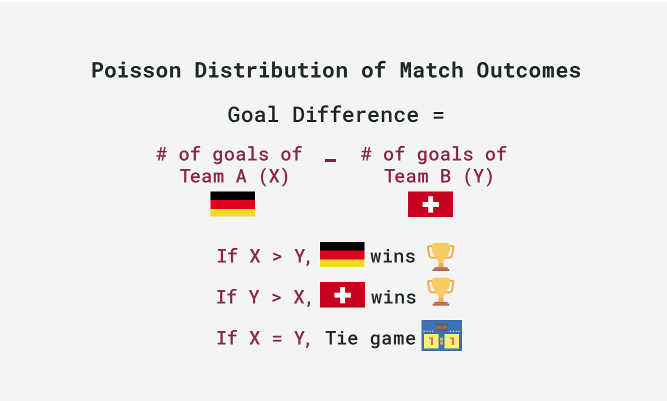
Consider a match between Germany and Switzerland. In the last two years, Switzerland conceded on average 0.60 goals per match. Germany played against Australia in the Confederations Cup, which ended in 3 to 2. Australia conceded on average 1.12 goals per match. Therefore, based on the number of goals Germany scored against Australia, we expect Germany to score $3 \times \frac{.60}{1.12} = 1.61$ goals against Switzerland.

We expect Germany to score fewer goals against Switzerland, because they have a better defense than Australia. Germany played 20 non-friendly matches in the two-year period, so we average the weighted scores over all opponents. After that, we run a similar analysis for Switzerland because the goal intensities are not symmetric. This means that the expected number of goals that Germany scores against Switzerland is not the same as the expected number of goals that Switzerland scores against Germany.

We can then model the number of goals team A scores against team B, using a Poisson distribution with the goal intensity $\lambda_{A,B}$ as its parameter.

The Poisson distribution is a distribution used to express the probability of a given number of events occuring in a fixed time period, given the average number of times the event occurs over that time period.

To know the outcome of a match, all we need is the goal difference, *Diff*= *X* - *Y*, where *X* and *Y* are the number of goals scored by teams A and B, respectively. If $X > Y$, the difference is positive and thus shows that team A wins; If $X = Y$, the difference is equal to zero and thus shows that the match ends in a tie; if $X < Y$, the difference is negative and thus shows that team B wins.



It turns out that, since both *X* and *Y* are Poisson distributed, *Diff* follows a Skellam distribution.

A Skellam distribution it is the discrete probability distribution of the difference X - Y of two independent random variables X and Y, each Poisson-distributed with respective expected values Lambda and Mu.

The Skellam distribution makes it simple to compute the probability of the aforementioned events. When ties are not allowed, as is the case in the knockout stage of the World Cup™, we need to evaluate the probabilities of winning in regular time, overtime (treated as a 30 minute independent match), and in a penalty shoot-out.

Monte Carlo Simulations

The World Cup™ is composed of a Group Stage, in which eight mini round-robin tournaments are held. Thereafter, the winners and runners-up compete in an elimination round called the knockout stage. We run hundreds of Monte Carlo simulations to obtain different scenarios, because there is a lot of uncertainty in the results of the Group Stage.

A Monte Carlo simulation, or probability simulation, is a technique used to understand the impact of risk and uncertainty in forecasting models.

- First, we simulate the final score for each of the 48 matches to know which countries advance to the next stage.
- Second, we cascade the probability of each country advancing through the knockout tree (quarter-finals, semi-finals, final, winner), using recursion.
- Finally, we average the probabilities of all the scenarios to obtain unconditional probabilities.

This model allows us to update the probabilities with actual, real-time match outcomes as they become available. We also use these match outcomes to augment the dataset when estimating goal intensities. There's no need to run simulations once we know who the round of 16 contenders will be.

Algorithms in Action

We used statistic probability to compute each country's chance of advancing through the tournament and ultimately winning the World Cup™. Here's how Machine Learning algorithms could be incorporated into our framework to achieve the outcome.

Machine Learning is often used for classification and regression tasks. Classification involves predicting a qualitative response, which takes on values in one of *K* different categories. Regression focuses on predicting a quantitative outcome. We could certainly use classification to predict the probabilities of the different match outcomes (namely win, draw, and lose), but it would be difficult to relate these to the winners and runners-up of the Group Stage. This is because the winners are derived from the number of points earned, the goal difference and the total number of goals scored. Instead, we should aim at learning $\lambda_{A,B}$ the expected number of goals team A will score against team B.

We can achieve this by using features that contain information on the structure of a team, its recent performance and renown, and even economic factors of the country they represent. Examples of these variables might be:

- Average age of the players
- Number of players competing in the Champions League
- Confederation of both the team in question and its opponent
- FIFA™ rank
- Winning probabilities extracted from bookmaker odds
- GDP per capita, normalized by the worldwide average

These features can be collected from historical data on official matches, as well as from contextual statistics at the time of the encounters. The numeric variables can be represented as the difference between both teams, while information such as the confederation of each country should be encoded as separate variables. Lastly, notice that because we use the number of goals each team scores as the response variable, each match amounts to two different observations, one per team.

Machine Learning in Motion

In Machine Learning and statistics, there's no such thing as a free lunch. This "No Free Lunch" theorem states that no one algorithm works best for every problem, or no one method dominates all others over all possible datasets. This is especially true in predictive modeling.

While we cannot say for sure that Support Vector Machines work better than decision trees or vice versa, it is important to try algorithms that are appropriate for the problem or task at hand. Good data scientists should test multiple algorithms, while using a hold-out "test set" of data to evaluate performance and select the winner.

Regression Trees

One Machine Learning algorithm we could employ is regression trees. A regression tree attempts to find the correct answer by asking as few Yes-No questions as possible. Each question should significantly narrow down the remainder of possible answers. The set of questions is selected in such a way that the variance among all training cases is minimized. At the time of prediction, we answer all the questions based on the features of the new case. **The prediction is simply the average of the response values for the training cases that followed the same path in the tree.**

Random Forest

While trees are simple and useful for interpretation, they typically are not competitive in terms of prediction accuracy. To overcome this we can build a Random Forest—a method that produces multiple trees and then combines those predictions to reach consensus, since we cannot grow different trees from the exact same training cases, and generally it is expensive (and difficult) to get more data.

Instead, we can take repeated samples, with replacement, from the single training data set, and built a separate tree for each of them. In any given data set, some observations may appear more than once while others do not appear at all. Just as diversity of thought helps humans reach better consensus, Random Forests benefit from having decorrelated trees. One way of achieving this is to constrain the algorithm to only consider a subset of features at each split (i.e. question) in each tree. Finally, agreement is obtained by averaging the predicted values of all trees in the forest.

When building Random Forests, it is crucial to fine-tune the following hyperparameters:

- **Max features:** the type of questions you are allowed to ask at each split
- **Max depth:** the total number of sequential questions that you are allowed to ask
- **Forest size:** the number of trees that should be *grown,* or trained

We accomplish this by trying different configurations and comparing the resulting error on the *Out-of-Bag* observations. Specifically, we get a prediction for each training case by using the trees that did not include the training case in the sample data set (a.k.a. bag) and then choose the configuration that performs the best.

Although Random Forests are a powerful tool for prediction, other learning algorithms, such as Artificial Neural Networks, should also be considered.

Ultimately, the performance of the ML approach should stacked up against common benchmarks, such as bookmaker odds or even a simple classification rule that selects the country with a highest FIFA™ rank as the winner of a match. We recommend starting with a simple model, and then iteratively applying more sophisticated methods.

Putting Paul to Work

Predictive algorithms are only as good as their value or utility, so we put Paul to the test. We pitted Paul against 253 enthusiastic and savvy Wizeliners. We challenged them to beat Paul and they rose to the challenge.

Employees were able to place their votes on the teams they thought would win, or tie, in every match. We tracked this with a leaderboard to see which Wizeliners made the best predictions in real-time.

Gamification

Each correct prediction was granted 10 points. Employees also had an additional "wildcard" opportunity to guess the ultimate winner of the World Cup™ tournament, regardless of their standing or prior predictions. Wildcard votes were accepted until the start of Round Two in the Groups Stage and worth 30 extra points if correct.

Paul's prediction for each matchup became visible once individual votes were locked in and the match was in progress.

Tournament Highligts

• France beat Croatia in the final match

VS

•Belgium and England were eliminated in the semi-finals

VS VS

•Brazil was eliminated in the quarter-finals

VS

•Spain and Portugal were eliminated in the Round of 16

• In a stunning upset, Germany was eliminated in the Groups Stage. Germany was a popular contender, and was predicted to win the tournament by many analysts.

The Most-voted Countries During the Wildcard Round

1. Germany : **27 %**
2. Brazil : **22.3 %**
3. Spain : **16.7 %**
4. France : **8.4 %**
5. Mexico : **8.4 %**

Accuracy of the Wizeline Trend

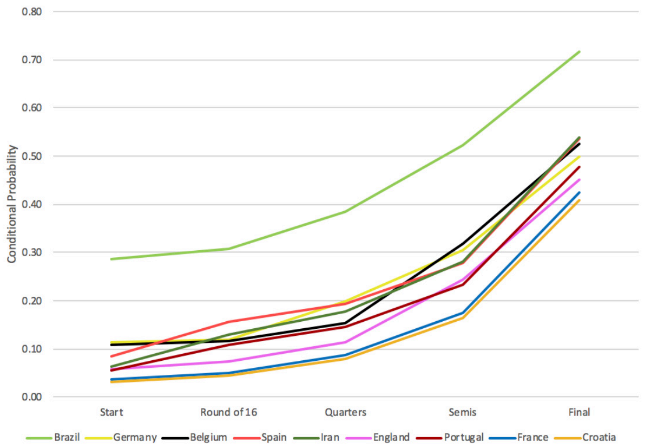
Computed from Wizeline employee participation

- Wizeliners correctly predicted 38 of the 64 match outcomes.
- Paul outscored 95 percent of individual Wizeliners.
- The Wizeline Trend, as an aggregate prediction, also outscored 95 percent of individual participants.

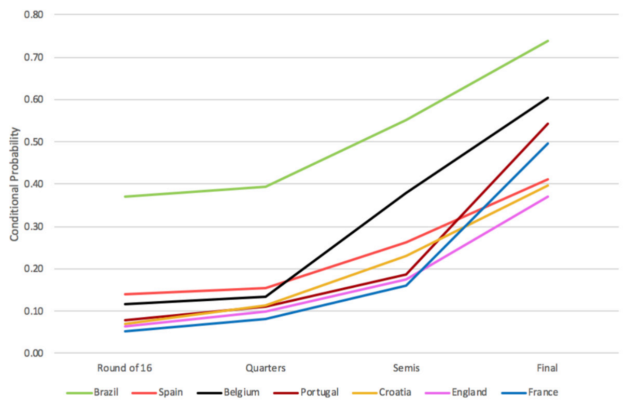
Insights from Conditional Probability Charts

Computed before the tournament begins and before the knockout stage begins

Probability of Winning, Conditioned on Reaching each Stage



Probability of Winning, Conditioned on Reaching each Stage



Insight #1: Iran was an outlier in the model. In 10 matches, Iran had "clean sheets" in nine matches (meaning it prevented its opponents from scoring any goals throughout an entire match) and allowed two goals in the 10th match. Therefore, Paul believed Iran's defense ability was strong. This is inaccurate once we consider Iran plays in a less competitive confederation.

Insight #2: In the second chart, we can see that the countries positioned on the left side of the draw had higher chances of winning the tournament.

Applications Beyond the World Cup

Wizeline works with clients who are at different stages of their digital strategy or transformation journey. We want to enable companies to exploit their data and turn it into valuable business insight.

We understand AI, and can help our clients build the right AI Strategy, using the right approach—whether that's Machine Learning, probability, data science, or crowdsourcing.

More importantly, we don't shy away from messy data. Our data scientists have experience building practical applications of data science, selecting an approach or model based on expected business outcomes.

When should teams use a certain approach? How should enterprises analyze the conversations coming from their audience? Wizeline helps enterprises figure out how Machine Learning and data science can add value to the business, reduce its spend, and optimize the areas that perform best.