

No Free Hunch (<http://blog.kaggle.com/>)

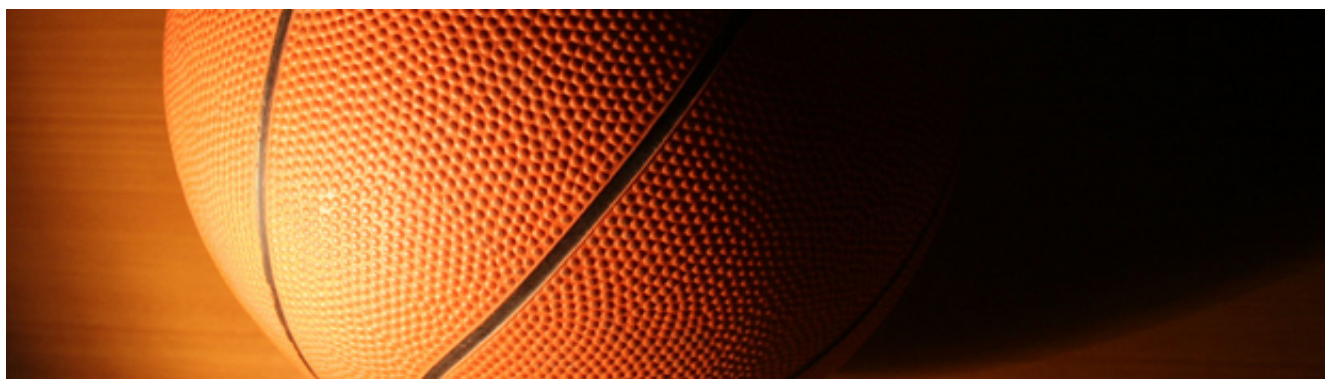


🏠 ([HTTP://BLOG.KAGGLE.COM](http://blog.kaggle.com/)) > MARCH MACHINE LEARNING MANIA 2016, WINNER'S INTERVIEW: 1ST PLACE, MIGUEL ALOMAR

⬅️ ([HTTP://BLOG.KAGGLE.COM/2016/05/13/BNP-PARIBAS-CARDIF-CLAIMS-MANAGEMENT-WINNERS-INTERVIEW-1ST-PLACE-TEAM-DEXTERS-LAB-DARIUS-DAVUT-SONG/](http://blog.kaggle.com/2016/05/13/BNP-PARIBAS-CARDIF-CLAIMS-MANAGEMENT-WINNERS-INTERVIEW-1ST-PLACE-TEAM-DEXTERS-LAB-DARIUS-DAVUT-SONG/)) ➡️ ([HTTP://BLOG.KAGGLE.COM/2016/05/06/MARCH-APRIL-2016-SCRIPTS-OF-THE-WEEK/](http://blog.kaggle.com/2016/05/06/MARCH-APRIL-2016-SCRIPTS-OF-THE-WEEK/))

March Machine Learning Mania 2016, Winner's Interview: 1st Place, Miguel Alomar

[Kaggle Team \(http://blog.kaggle.com/author/kaggleteam/\)](http://blog.kaggle.com/author/kaggleteam/) | 05.10.2016



The annual [March Machine Learning Mania \(https://www.kaggle.com/c/march-machine-learning-mania-2016\)](https://www.kaggle.com/c/march-machine-learning-mania-2016) competition sponsored by [SAP \(https://www.kaggle.com/c/march-machine-learning-mania-2016/details/about-the-sponsor\)](https://www.kaggle.com/c/march-machine-learning-mania-2016/details/about-the-sponsor) challenged Kagglers to predict the outcomes of every possible match-up in the 2016 men's NCAA basketball tournament. Nearly 600 teams competed, but only the first place forecasts were robust enough against upsets to top this year's bracket. In this blog post, Miguel Alomar describes how calculating the offensive and defensive efficiency played into his winning strategy.

The Basics

What was your background prior to entering this challenge?

I earned a Master's Degree in Computer Science from UIB in Mallorca, Spain. For nearly 20 years, I have been involved in software development, business intelligence and data warehousing. Recently, I have developed an interest in analytics and forecasting.

1

(<http://l>

[machin](#)

[learning](#)

[mania-](#)

[2016-](#)

[winners](#)

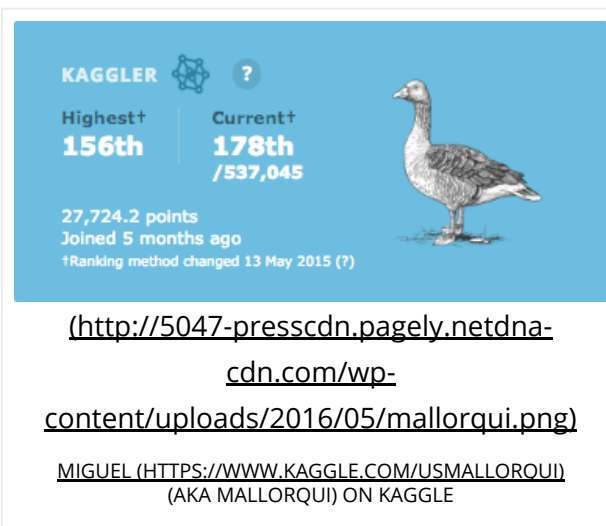
[interview](#)

[1st-](#)

[place-](#)

[miguel-](#)

[alomar/](#)



Do you have any prior experience or domain knowledge that helped you succeed in this competition?

In Spain, I played amateur basketball for 10 years. I like to think that is the reason I won.

The truth is I missed most of the basketball games this season and did not have a good feel for the any of the team's quality. That most likely helped me because if I had seen more games, my judgment may have changed some of the forecasts. Normally, I am pretty bad at picking winners.

How did you get started competing on Kaggle?

I found Kaggle through some data science lessons I was taking on [Coursera](https://www.coursera.org/learn/machine-learning) (<https://www.coursera.org/learn/machine-learning>).

What made you decide to enter this competition?

I really like analytics and sports so I thought it was a perfect competition for me.

But the key factor is that moderators and other members make it easy to enter, they provide lots of help, data, advice and feedback. Data is already formatted and prepared so the data gathering and manipulating task is made very easy. Some members of the community seemed more interested in sharing and discovering new methods and insights than in winning the competition.

Let's get technical

What preprocessing and supervised learning methods did you use?

I used logarithmic regression and random forests. I did try ADA Boost but didn't get very good results so I didn't use it in my final model.

What was your most important insight into the data?

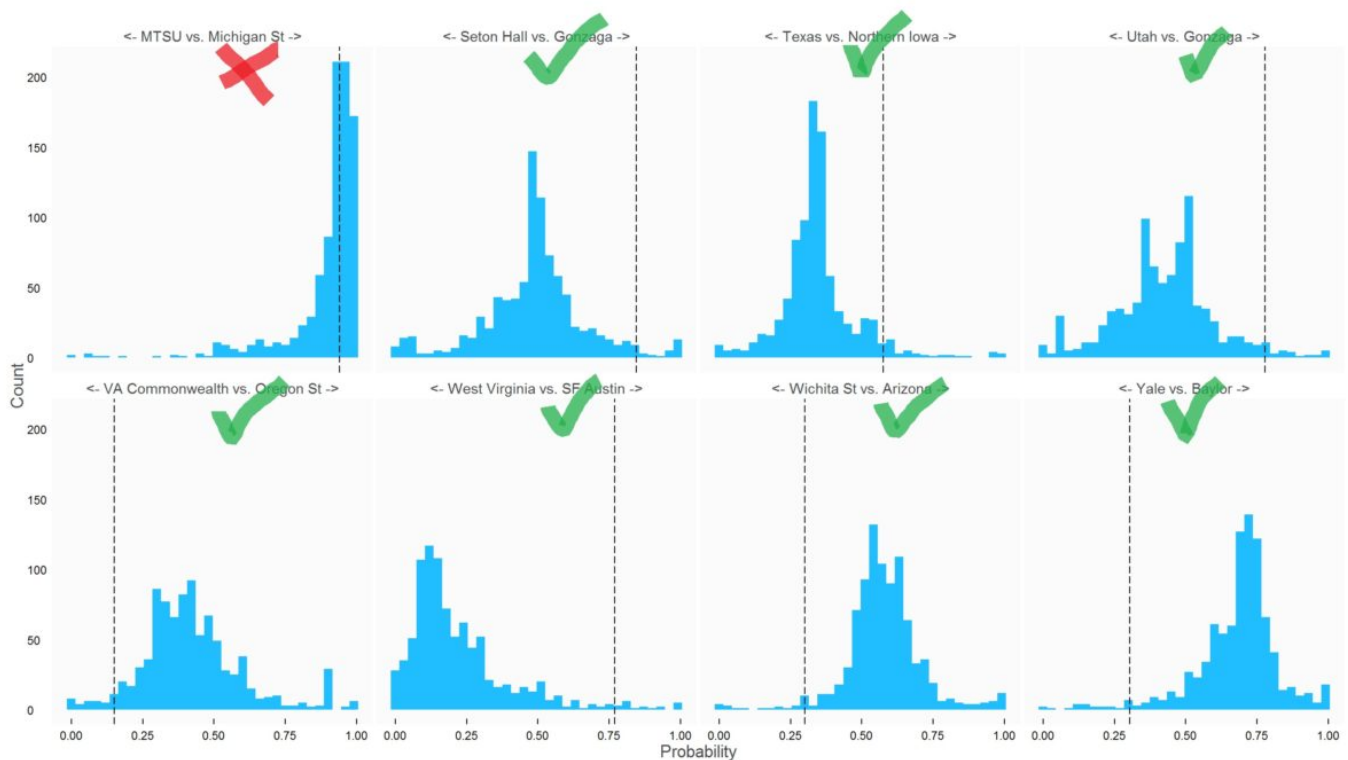
The data behind this competition is very simple, the box stats from basketball games are very simple to understand. The key factor for me was the offensive and defensive efficiency, how to calculate those? What weight to give to strength of schedule? Can you "penalize" a team because they haven't played against the best teams in the nation? Can you lower their rating for something that didn't happen?

Those are the kind of questions I was trying to answer, I developed several models with different degrees of adjusted efficiency ratings and checked their scores against past seasons.

Since my scores in Stage1 of the competition were not very good, I kept changing my model after Stage1 was closed.

My goal for next year is to formally test those different models to find out if there is any validity to my ideas.

kaggle.com - March Machine Learning Mania 2016 (Early rounds upsets)



http://5047-presscdn.pagely.netdna-cdn.com/wp-content/uploads/2016/05/first_round_upsets.jpg

Were you surprised by any of your findings?

After building the submission files, I put them into brackets using a script provided by one of the Kaggle members. My first model had a more conservative look to it and my second model (the final winner) just didn't look right to me. Teams like SF Austin, Indiana and Gonzaga were predicted to go very far in the bracket. I almost scrapped it but since it was my 2nd model I decided to go with it. This model got most of the first round upsets right, that surprised me.



(http://5047-presscdn.pagely.netdna-cdn.com/wp-content/uploads/2016/05/Bracket_manual2.jpg)

CLICK TO EXPAND.

Which tools did you use?

I used R, R studio and SQL.

How did you spend your time on this competition?

I would say my time allocation was 35% reading forums and blogs, 15% manipulating data, 25% building models and 25% evaluating results.

What was the run time for both training and prediction of your winning solution?

Five minutes. I trained my model using only 2016 data, so the amount of data to process is very small.

Bio

Miguel Alomar has a Master's Degree in Computer Science from UIB in Mallorca, Spain. For nearly 20 years, he has been involved in software development, business intelligence and data warehousing.

Sponsored

What's your dream travel experience?

Win up to a \$2,500 scholarship to study or volunteer in a country of your choice.



Enter today ▶

Revitalize Your Body & Improve Your Health!



O₂

2 Comments

No Free Hunch

1 Login ▾

Recommend 5 Share

Sort by Best ▾



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS ?



Name



Simon Sayz • 11 days ago

Nice post. Congratulations. I am starting on Kaggle and it's always nice to get infos on how others did it. Thanks.

^ | ▾ • Reply • Share ›



Anand Sharma • 2 years ago

Awesome!! and congratulations :yesss!!

^ | ▾ • Reply • Share ›

Omaha, NE's NEW RULE

Drivers With No Tickets In 3 Years Are In For A Big Surprise

[Learn More](#)

Sponsored by [Comparisons.org](#)



[Report ad](#)



<https://www.facebook.com/kaggle>



<https://twitter.com/kaggle>

