

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



⬆ ([HTTP://BLOG.KAGGLE.COM/](http://BLOG.KAGGLE.COM/)) > Q&A WITH GREGORY MATTHEWS AND MICHAEL LOPEZ, 1ST PLACE IN MARCH ML MANIA

⬅ ([HTTP://BLOG.KAGGLE.COM/2014/04/22/GE-ANNOUNCES-THE-WINNERS-OF-FLIGHT-QUEST-2/](http://BLOG.KAGGLE.COM/2014/04/22/GE-ANNOUNCES-THE-WINNERS-OF-FLIGHT-QUEST-2/)) ➡
([HTTP://BLOG.KAGGLE.COM/2014/04/18/WINNING-THE-GALAXY-CHALLENGE-WITH-CONVNETS/](http://BLOG.KAGGLE.COM/2014/04/18/WINNING-THE-GALAXY-CHALLENGE-WITH-CONVNETS/))

Q&A with Gregory Matthews and Michael Lopez, 1st Place in March ML Mania

[Kaggle Team \(http://blog.kaggle.com/author/kaggleteam/\)](http://blog.kaggle.com/author/kaggleteam/) | 04.21.2014

Gregory Matthews and Michael Lopez are the members of team [One shining MGF \(https://www.kaggle.com/c/march-machine-learning-mania/leaderboard/private\)](https://www.kaggle.com/c/march-machine-learning-mania/leaderboard/private) who climbed up to first place during a raucous ride on the leaderboard of Kaggle's March Machine Learning Mania. After all predictive models were frozen on March 19, things unfolded to real-world game results in the 2014 NCAA Tournament [see the other blog posts tagged as [march-mania \(http://blog.kaggle.com/tag/march-mania/\)](http://blog.kaggle.com/tag/march-mania/)]. We asked Greg and Mike to tell us how they approached the problem, working together for the first time on Kaggle.

What was your background prior to entering this challenge?

Greg: I have a Ph.D. in statistics from the University of Connecticut, completed a post-doc at the University of Massachusetts-Amherst, and in the fall will be an Assistant Professor of statistics at Loyola University Chicago. This was my first Kaggle contest.

Mike: I am a Ph.D. candidate in biostatistics from Brown University, and in the fall will be an Assistant Professor of statistics at Skidmore College. This was also my first Kaggle contest.

What made you decide to enter?

Greg: I've always been interested in evaluating and predicting sports using statistical methods. I am particularly interested in professional football, professional baseball, and college basketball. This contest seemed right up my alley.

Mike: Greg emailed me.

What preprocessing and supervised learning methods did you use?

Greg & Mike: Our winning submission was the combination of two models, a margin-of-victory based model (MOV) and an efficiency model using Ken Pomeroy's data (KP). For first round games, the MOV model used the actual spread posted in Las Vegas for each game; for future games, we used previous

game outcomes to predict a margin of victory. At the end, the spread (or the expected margin) was converged into a probability using logistic regression. For the KP model, we tried different regression models using different team-wide efficiency metrics, eventually settling one that minimized our loss function on the training data. At the end, we used a weighted average of the two probabilities (one from the MOV model, one from the KP model) as our final submission.

What was your most important insight into the data?

Greg: The Las Vegas line is absolutely incredible at predicting games. As they say, if you can't beat them, use their data in a Kaggle contest. Also, when training the models, we didn't just try to predict the tournament games, as there is a relatively small number of those types of games. Instead, we also trained our models on regular season data, too.

Mike: Like Greg said, it seemed silly to only train our models on a sample of 63 tournament games each season, when, in fact, there are hundreds of games played each week. Not sure it helped us, but we also ignored tournament specific information (i.e. a team's seed number).

Were you surprised by any of your insights?

Greg: I was surprised by how well our simple models performed. Using the right data was MUCH more important to our models performing well than using more sophisticated models.

Mike: I think we gave ourselves a chance with a good model, but there was probably a decent amount of luck involved, too. Also, identifying the specific loss function for this Kaggle contest, and where it comes from, seemed to help our model.

Which tools did you use?

Greg & Mike: R and R Studio

What have you taken away from this competition?

Greg: Don't cheat.

Mike: I should always say yes when Greg emails me.

<http://5047-presscdn.pagely.netdna-cdn.com/wp-content/uploads/2014/02/mania.jpg>

Michael Lopez is a 4th-year Ph.D. student in the Department of Biostatistics at Brown University. Mike's website is statsbylopez.com (<http://statsbylopez.com/>)

Gregory J. Matthews is currently a lecturer and the Associate Director of the Institute for Computational Biology, Biostatistics, and Bioinformatics in the Department of Public Health at UMass-Amherst. Greg's website is statsinthewild.com (<http://statsinthewild.com/>)



[MARCH MANIA \(HTTP://BLOG.KAGGLE.COM/TAG/MARCH-MANIA/\)](http://blog.kaggle.com/tag/march-mania/)



BUFFERED VPN

Customer Service like no other VPN

GET STARTED
30-DAY MONEY-BACK

Voted as No. 1 on BestVPN.com

What's your dream travel experience?

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

Scholar TRIPS INTERNATIONAL TRAVEL SCHOLARSHIPS

Enter today

0 Comments


No Free Hunch

1 Login

Recommend

Share





Sort by Best



Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS



Be the first to comment.

Bitcoin is Dead - This Will Make Investors Rich in 2018

If you suspect Bitcoin is going to crash, I just want you to know, you're right. Here is the truth about Bitcoin that no one else will tell you.

[Learn More](#)

Sponsored by **Bonner and Partners**



[Report ad](#)



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



<https://twitter.com/kaggle>

