

Theran Meadows
Project 1: Master Chef
Milestone 3
June 30, 2024

Business Problem

The question/problem I am trying to answer is whether or not the winner of each season was the most skilled contestant of their respective opponents for that season to be crowned the winner. I would like to implement my own point system to confirm or deny my theory and provide a mock recommendation on how the FOX tv network could change how they decide who wins the Master chef reward.

I would also like to show a predictive classification model to see how the show performs between its own method and my own

Background/History

Currently the show has had a consistent number of contestants per season at 20. The show will then filter through one or two challenges per episode usually eliminating at least one person but sometimes two or more can also be eliminated. There may be partner challenges or group team challenges as well. The single or team winner of a challenge will then be given immunity from being eliminated and will therefore usually sit the following challenge out. The final 3 are then considered the finalists and complete through 3 final rounds cooking. The first round is the appetizer, the second is the entree and the third is the dessert. After the appetizer or the entree is usually when one of the finalists is eliminated. The winner is then chosen after both have completed the challenge

Data Explanation (Data Prep/Data Dictionary/etc)

The data is from the Master Chef Wiki Fan page that keep track of the contestants and the outcome of each challenge. It does keep track of the wins, tops, bottoms, and safes that a person landed in the current challenge but does not keep a running total. In my dashboard and predictive model, I used my own point system to keep a running total for every contestant. I used Excel and Sql to prepare the data and Python for the predictive model.

Methods

The target value for any of the predictive model tests was the winner variable. The purpose of the model was to see if the default Master Chef process was accurate or if another model should be used. A decision tree was used to compare results between the current method and my recommended point system method.

Analysis

After running the decision tree for the current method of elimination and consequently, choosing a winner, I found that it was only accurate about 15% of the time. Using my method, I revised the CSV file with all the contestant stats to where the contestant with the most points would be the winner along with the 2nd, 3rd, 4th and so on places even if they were different in the current method. I ran the model again to see if my method would be accurate in declaring the winner. Unfortunately, I found that it was only a 2% increase in accuracy for finding who the winner would be. This was a little disappointing as it would likely mean that some of the contestants I felt should not

win might still end up being the winner or at least in the top when I thought they would not even be in the top three. It's possible that the attributes I chose were simply not enough or that there wasn't enough data. After all, with all the contestants measured, it is still under 300 rows.

If I were to test again, I would want to add more attributes like cooking background, hobbies, professions and personality traits. I would also want data to be collected every episode on each contestant with their behaviors, choice dishes, and judges feedback. I think these extra variables might also help determine who would be close to winning.

Conclusion

In conclusion, Master Chef currently has an inconsistent way of choosing an eventual winner of a 20 contestant season. In my own research and implementation of a point system, I have found that over 50% of all seasons reward those who do not win the challenges often and that those who do win have a greater likelihood of going home. From a predictive model standpoint, the current method had about 15% while the point system method had only 17%. More data attributes are needed to possibly increase the accuracy of both models.

Assumptions/ limitations

The data set, although complete as far as having every episode and contestant evaluated up till the 10th season, the data set is still under 300 rows. There may not be enough data to have a precise accuracy measurement.

Challenges

Because of the small data set, my model runs have run into over fitting a lot of times with accuracy being 100%.

Future Uses/Additional Applications

This data set will be growing as even now there is a current season of Mast Chef airing. This set can always be added to but I think there won't be enough data fast enough to predict who would win.

Analyzing the quality of characteristics of contestants as well could also be used to predict who would win.

Recommendations

My recommendation would be to implement that point system into the show to see who has the most wins at the end of a season rather than reward those that happen to cook good enough to pass a challenge but not be in the top cooks.

My recommendation is also to continue the analysis on discovering new data; attributes, behaviors, background, etc... to also use in future predictions model tests.

Implementation Plan

This system would be used going forward for future seasons. It would have to be established to all contestants, judges and audience to have all people aware of how the show and competition will work.

Ethical Assessment

There are no major ethical issues with the data set now as it is all public information. Nonetheless, going forward, judges in the show would have to not be biased towards any one or group of contestants. Telling someone to not be biased is impossible though. So no matter what, there will be an element of bias as long as judges have a say in who wins.

10 Q&A

- 1) Which contestant has the most wins overall from all seasons?

name	status	num_wins
abc Filter...	abc Filter...	abc Filter...
Natasha Crnjac	2	13
Elizabeth Cauvel	2	12
Caitlin "Cate" Meade	3	11

This is a screenshot of the data uploaded into a Postgresql server using the following query

```
select name, status, num_wins FROM mc_contestants
join mc_stats on mc_contestants.contestant_id = mc_stats.contestant_id
order by num_wins DESC
limit 3;
```

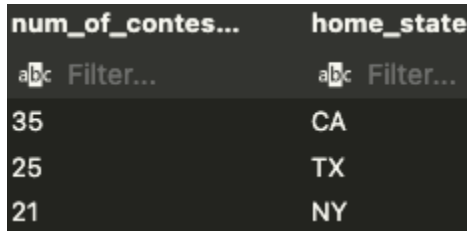
Notice how these top 3 contestants with the most points did not come in 1st place.

- 2) How many contestants scored 0 points or lower?

The following query will show that there are 54 contestants of all the seasons that did not score any points.

```
select count(*) FROM mc_contestants
join mc_stats on mc_contestants.contestant_id = mc_stats.contestant_id
where num_wins = 0 or num_wins < 0
```

3) Where are the majority of contestants coming from?



The screenshot shows a table with two columns: 'num_of_contes...' and 'home_state'. The table has three rows of data. The first row shows 35 contestants from CA (California). The second row shows 25 contestants from TX (Texas). The third row shows 21 contestants from NY (New York). Above the table, there are two search filters labeled 'abc Filter...'.

num_of_contes...	home_state
35	CA
25	TX
21	NY

The following query gives us the above screenshot of the number of contestants that came from the top 3 states where contestants are from. With every episode filmed in California with the exception of travel episodes, it's no wonder that the most contestants come from California.

4) What is the average age among the winners of the aired seasons so far?

The following query will show that 31 is the average age among the winners only.

```
select round(avg(age)) as AVG_age_of_winners FROM mc_contestants  
join mc_stats on mc_contestants.contestant_id = mc_stats.contestant_id  
where status = 1
```

5) Are there other factors in a contestant that are being measured besides wins and losses
Etc?

For the models tested, gender, home state, and age were also considered besides the stats of the contestants. (wins, tops, safe, bottoms)

For further analysis, I would definitely consider cooking background, hobbies, professions, personality traits to be predictors of winners.

6) How would the point system work?

The point system works by giving a rolling point total to all contestants every challenge. For contestants that were in the top dishes of a challenge, usually there are 3 chosen, 1 point is given. 1 winner is chosen from the top dishes to which that contestant will get an additional point. On the contrary, usually 3 dishes are also chosen as the worst of the challenge to which 1

point is deducted. Of those 3 bottom dishes, 1 contestant is eliminated and they would no longer continue in the challenges or gathering points.

It should also be noted that more than one contestant can be eliminated at the end of a challenge. Some episodes also have team challenges where two teams are made and one captain per team is chosen to lead. Whichever team loses will usually have the captain be up for elimination unless there was someone on the losing team that was a standout for their poor performance. Team challenges can also consist of partner challenges between two people. The losing team can have one or both of the contestants leave.

7) How can you validate the model's accuracy?

Validating can come from using other predictive models to see if the accuracy is around the same percentage.

8) How could this impact the audience/ viewers of the show?

If this point system were used, the impact has more benefits than risks. The reason being is that since the first season to the present season, viewership went from 5.2 million to 2.2 million. The show is at a point where doing something 'different' might be useful.

Another benefit is that the most deserving cooks would actually have a chance at winning. Unfortunately, that would also mean less dramatic moments on television.

9) What do similar shows use for determining who gets eliminated?

Top Chef also works with eliminating chefs that had poor performance but have a 'Last Chance' challenge to be in the running to win again. Chopped also works that same way which is why a 'new change' might help Masterchef get more views.

10) How soon can a prediction be accurately made?

It's hard to say. The closer you get to the final challenges, you already would have a 33% chance of choosing the winner. The further away you are, you have more choices but as supported by evidence from the presentation, some contestants can have multiple dishes that were in the bottom but still end up winning.

Adding more qualities to the background of a contestant would also help to see where the trajectory could lead them along with data gathered on each episode.

References

Wikimedia Foundation. (2024, June 13). *MasterChef (American TV series)*.

Wikipedia. [https://en.wikipedia.org/wiki/MasterChef_\(American_TV_series\)](https://en.wikipedia.org/wiki/MasterChef_(American_TV_series))