Austin Lowey

Predicting Survivor Winners with Machine Learning

LinkedIn: Linkedin.com/in/Austin-Lowey

Portfolio: AustinLowey.github.io

GitHub Pages

Predicting Survivor Winners with Machine Learning

Leveraged supervised learning techniques to predict the winner of the competition TV show, Survivor, based on historical performance data.

Part 1 - Data Engineering

Data ETL to build SQL database of contestant features for 45 Survivor seasons

Part 2 - Machine Learning

Trained multiple ML models using my database to predict contestant placements

Tools Used:

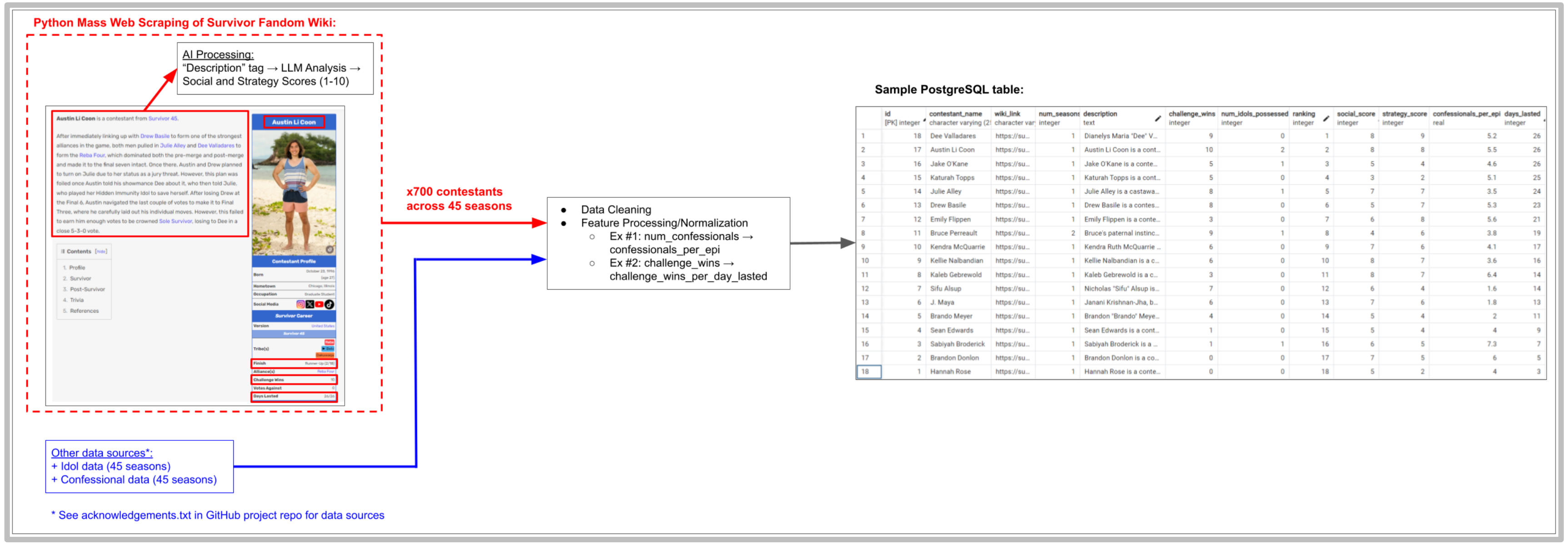
Python: scikit-learn, pandas, NumPy, psycopg2, Beautiful Soup/requests, OpenAI LLM API

Databases: PostgreSQL, pgAdmin

Part 1: Data Engineering

Populated a database with contestant data features for ML modeling.

- Planned ML model features and researched potential data sources
- Data ETL, including mass web scraping/parsing of the Survivor Wiki
- Integrated OpenAI's LLM API to conduct automated AI-analysis on contestant descriptions, generating social and strategy scores on pre-defined criteria

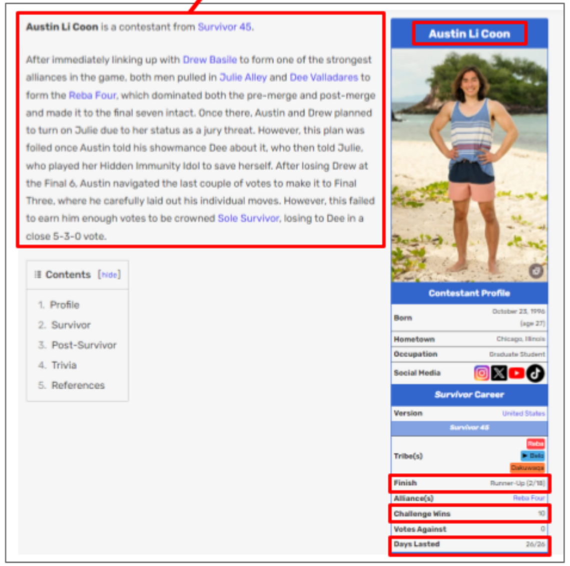


	id [PK] integer	contestant_name character varying (2	wiki_link character var	num_seasons integer	description text	challenge_wins integer	num_idols_possessed integer	ranking integer	social_score integer	strategy_score integer	confessionals_per_epi real	days_lasted integer
1	18	Dee Valladares	https://su...	1	Dianelys Maria "Dee" V...	9	0	1	8	9	5.2	26
2	17	Austin Li Coon	https://su...	1	Austin Li Coon is a cont...	10	2	2	8	8	5.5	26
3	16	Jake O'Kane	https://su...	1	Jake O'Kane is a conte...	5	1	3	5	4	4.6	26
4	15	Katurah Topps	https://su...	1	Katurah Topps is a cont...	5	0	4	3	2	5.1	25
5	14	Julie Alley	https://su...	1	Julie Alley is a castawa...	8	1	5	7	7	3.5	24
6	13	Drew Basile	https://su...	1	Drew Basile is a contes...	8	0	6	5	7	5.3	23
7	12	Emily Flippen	https://su...	1	Emily Flippen is a conte...	3	0	7	6	8	5.6	21
8	11	Bruce Perreault	https://su...	2	Bruce's paternal instinc...	9	1	8	4	6	3.8	19
9	10	Kendra McQuarrie	https://su...	1	Kendra Ruth McQuarrie ...	6	0	9	7	6	4.1	17
10	9	Kellie Nalbandian	https://su...	1	Kellie Nalbandian is a c...	6	0	10	8	7	3.6	16
11	8	Kaleb Gebrewold	https://su...	1	Kaleb Gebrewold is a c...	3	0	11	8	7	6.4	14
12	7	Sifu Alsup	https://su...	1	Nicholas "Sifu" Alsup is...	7	0	12	6	4	1.6	14
13	6	J. Maya	https://su...	1	Janani Krishnan-Jha, b...	6	0	13	7	6	1.8	13
14	5	Brando Meyer	https://su...	1	Brandon "Brando" Meye...	4	0	14	5	4	2	11
15	4	Sean Edwards	https://su...	1	Sean Edwards is a cont...	1	0	15	5	4	4	9
16	3	Sabiyah Broderick	https://su...	1	Sabiyah Broderick is a ...	1	1	16	6	5	7.3	7
17	2	Brandon Donlon	https://su...	1	Brandon Donlon is a co...	0	0	17	7	5	6	5
18	1	Hannah Rose	https://su...	1	Hannah Rose is a conte...	0	0	18	5	2	4	3

Part 1: Data Engineering

Python Mass Web Scraping of Survivor Fandom Wiki:

<u>AI Processing:</u>
"Description" tag → LLM Analysis → Social and Strategy Scores (1-10)



Austin Li Coon is a contestant from Survivor 45.

After immediately linking up with Drew Basile to form one of the strongest alliances in the game, both men pulled in Julie Alley and Dee Valladares to form the Reba Four, which dominated both the pre-merge and post-merge and made it to the final seven intact. Once there, Austin and Drew planned to turn on Julie due to her status as a jury threat. However, this plan was foiled once Austin told his showmance Dee about it, who then told Julie, who played her Hidden Immunity Idol to save herself. After losing Drew at the Final 6, Austin navigated the last couple of votes to make it to Final Three, where he carefully laid out his individual moves. However, this failed to earn him enough votes to be crowned Sole Survivor, losing to Dee in a close 5-3-0 vote.

Contents [hide]

1. Profile
2. Survivor
3. Post-Survivor
4. Trivia
5. References

Austin Li Coon

Contestant Profile

Born	October 23, 1996 (age 27)
Hometown	Chicago, Illinois
Occupation	Graduate Student
Social Media	

Survivor Career

Version	United States
Survivor 45	
Tribe(s)	Reba, Belo, Dakuwaqa
Finish	Runner-Up (2/18)
Alliance(s)	Reba Four
Challenge Wins	10
Votes Against	0
Days Lasted	26/26

x700 contestants across 45 seasons

- Data Cleaning
- Feature Processing/Normalization
  - Ex #1: num_confessionals → confessionals_per_epi
  - Ex #2: challenge_wins → challenge_wins_per_day_lasted

<u>Other data sources*:</u>
+ Idol data (45 seasons)
+ Confessional data (45 seasons)

* See acknowledgements.txt in GitHub project repo for data sources

Sample PostgreSQL table:

contestant_name character varying (2!	wiki_link character var	num_seasons integer	description text	challenge_wins integer	num_idols_possessed integer	ranking integer	social_score integer	strategy_score integer	confessionals_per_epi real	days_lasted integer
Dee Valladares	https://su...	1	Dianelys Maria "Dee" V...	9	0	1	8	9	5.2	26
Austin Li Coon	https://su...	1	Austin Li Coon is a cont...	10	2	2	8	8	5.5	26
Jake O'Kane	https://su...	1	Jake O'Kane is a conte...	5	1	3	5	4	4.6	26
Katurah Topps	https://su...	1	Katurah Topps is a cont...	5	0	4	3	2	5.1	25
Julie Alley	https://su...	1	Julie Alley is a castawa...	8	1	5	7	7	3.5	24
Drew Basile	https://su...	1	Drew Basile is a contes...	8	0	6	5	7	5.3	23
Emily Flippen	https://su...	1	Emily Flippen is a conte...	3	0	7	6	8	5.6	21
Bruce Perreault	https://su...	2	Bruce's paternal instinc...	9	1	8	4	6	3.8	19
Kendra McQuarrie	https://su...	1	Kendra Ruth McQuarrie ...	6	0	9	7	6	4.1	17
Kellie Nalbandian	https://su...	1	Kellie Nalbandian is a c...	6	0	10	8	7	3.6	16
Kaleb Gebrewold	https://su...	1	Kaleb Gebrewold is a c...	3	0	11	8	7	6.4	14
Sifu Alsup	https://su...	1	Nicholas "Sifu" Alsup is...	7	0	12	6	4	1.6	14
J. Maya	https://su...	1	Janani Krishnan-Jha, b...	6	0	13	7	6	1.8	13
Brando Meyer	https://su...	1	Brandon "Brando" Meye...	4	0	14	5	4	2	11
Sean Edwards	https://su...	1	Sean Edwards is a cont...	1	0	15	5	4	4	9
Sabiyah Broderick	https://su...	1	Sabiyah Broderick is a ...	1	1	16	6	5	7.3	7
Brandon Donlon	https://su...	1	Brandon Donlon is a co...	0	0	17	7	5	6	5
Hannah Rose	https://su...	1	Hannah Rose is a conte...	0	0	18	5	2	4	3

Part 2: Machine Learning

Explored multiple ML and mathematical modeling techniques, with Random Forest Regression ML model yielding great results.

- Data split into 60/20/20% Train/Validation/Test
- Mathematical modeling: Continuous probability transformation function of contestant placement – Linear vs. Exponential Decay

- Linear Regression baseline model (v1)
- Random Forest Regression successor model (v2)
  - Good at handling noisy data
  - Captures non-linear feature relationships
  - Feature importance values useful for model interpretability

Expected ideal α range

rank	P(α=0.5)	P(α=0.4)	P(α=0.3)	P(α=0.2)	P(α=0.1)
1	1	1.00	1.00	1.00	1.00
2	0.61	0.67	0.74	0.82	0.90
3	0.37	0.45	0.55	0.67	0.82
4	0.22	0.30	0.41	0.55	0.74
5	0.14	0.20	0.30	0.45	0.67
6	0.08	0.14	0.22	0.37	0.61
7	0.05	0.09	0.17	0.30	0.55
8	0.03	0.06	0.12	0.25	0.50
9	0.02	0.04	0.09	0.20	0.45
10	0.01	0.03	0.07	0.17	0.41
11	0.01	0.02	0.05	0.14	0.37
12	0.00	0.01	0.04	0.11	0.33
13	0.00	0.01	0.03	0.09	0.30
14	0.00	0.01	0.02	0.07	0.27
15	0.00	0.00	0.01	0.06	0.25
16	0.00	0.00	0.01	0.05	0.22
17	0.00	0.00	0.01	0.04	0.20
18	0.00	0.00	0.01	0.03	0.18
19	0.00	0.00	0.00	0.03	0.17
20	0.00	0.00	0.00	0.02	0.15

Desirable weighting spread for top players

Decay is appropriate for N=16 through N=20 (range for number of contestants for each season)

Model Performance: Test Dataset

- 29% of placement predictions were exact or off by 1.
- 57% were within 3 places.

Model Predictions on Test Dataset (25/165 rows; full data on GitHub)

contestant_name	actual_probability_from_placement	predicted_probability	placement	predicted_placement	season_total_contestants	season
Kelly Wiglesworth	0.82	0.56	2	4	16	1
Rudy Boesch	0.67	0.55	3	4	16	1
Sean Kenniff	0.45	0.44	5	5	16	1
Dirk Been	0.11	0.13	12	11	16	1
Colby Donaldson	0.82	0.50	2	4	16	2
Amber Brkich	0.37	0.28	6	7	16	2
Jerri Manthey	0.25	0.26	8	8	16	2
Kimmi Kappenberg	0.11	0.40	12	6	16	2
Kelly Goldsmith	0.20	0.52	9	4	16	3
Diane Ogden	0.05	0.05	16	16	16	3
Rob Mariano	0.17	0.70	10	3	16	4
Patricia Jackson	0.06	0.05	15	16	16	4
Peter Harkey	0.05	0.05	16	16	16	4
Helen Glover	0.55	0.36	4	6	16	5
Penny Ramsey	0.30	0.18	7	10	16	5
Erin Collins	0.20	0.28	9	7	16	5
Robb Zbacnik	0.14	0.07	11	14	16	5
Stephanie Dill	0.11	0.13	12	11	16	5
Tanya Vance	0.06	0.11	15	12	16	5
John Raymond	0.05	0.04	16	16	16	5
Rob Cesternino	0.67	0.60	3	4	16	6
Daniel Lue	0.07	0.16	14	10	16	6
Janet Koth	0.06	0.08	15	14	16	6
Sandra Diaz-Twine	1.00	0.54	1	4	16	7
Jon Dalton	0.67	0.49	3	5	16	7

Model Performance: S46 Week-to-Week Predictions

- Consistently forecasted top contestants, especially the top 2.
- Predicted Charlie (<u>very</u> close 2nd place) as 1st place winner as early as Week 5, and never predicted him to place higher than 3rd.

Contestant	Actual Placement	Week 4		Week 5		Week 6		Week 7		Week 8		Week 9		Week 10		Week 11		Week 12		Actual Placement
		Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	Predicted Placement	Predicted Win Probability	
Kenzie Petty	1	5	8.4%	6	6.6%	5	9.6%	3	12.9%	2	17.0%	2	21.9%	2	19.1%	2	20.8%	2	20.8%	1
Charlie Davis	2	3	12.4%	1	15.3%	3	13.5%	2	14.8%	1	23.8%	1	23.8%	1	24.6%	1	28.3%	1	40.0%	2
Ben Katzman	3	8	5.3%	8	5.9%	7	7.5%	6	9.8%	4	11.7%	4	12.0%	4	14.6%	3	20.6%	4	14.2%	3
Liz Wilcox	4	12	3.5%	10	3.8%	10	3.1%	9	4.6%	8	4.2%	6	8.3%	6	5.5%	6	5.1%	5	6.7%	4
Maria Shrime Gonzalez	5	6	8.1%	5	8.7%	8	6.9%	4	10.9%	6	8.5%	5	10.4%	3	18.7%	4	17.5%	3	18.3%	5
Q Burdette	6	2	12.6%	2	15.2%	1	16.2%	5	10.2%	7	4.7%	7	4.2%	5	14.2%	5	7.7%	6	0%	6
Venus Vafa	7	14	2.1%	11	3.5%	12	2.0%	10	2.4%	9	2.6%	8	3.00%	7	3.3%	7	0%	7	0%	7
Tiffany Nicole Ervin	8	1	15.0%	3	13.6%	2	15.0%	1	17.4%	3	16.6%	3	16.4%	8	0%	8	0%	8	0%	8
Hunter McKnight	9	10	3.8%	7	6.1%	4	10.7%	7	9.1%	5	11.0%	9	0%	9	0%	9	0%	9	0%	9
Tevin Davis	10	4	9.4%	4	11.0%	6	9.3%	8	8.0%	10	0%	10	0%	10	0%	10	0%	10	0%	10
Soda Thompson	11	9	5.1%	9	4.1%	9	3.5%	11	0%	11	0%	11	0%	11	0%	11	0%	11	0%	11
Tim Spicer	12	13	3.3%	13	3.0%	11	2.9%	12	0%	12	0%	12	0%	12	0%	12	0%	12	0%	12
Moriah Gaynor	13	11	3.6%	12	3.1%	13	0%	13	0%	13	0%	13	0%	13	0%	13	0%	13	0%	13
Jemila Hussain-Adams	14	7	7.6%	14	0%	14	0%	14	0%	14	0%	14	0%	14	0%	14	0%	14	0%	14
Bhanu Gopal	15	15	0%	15	0%	15	0%	15	0%	15	0%	15	0%	15	0%	15	0%	15	0%	15
Randen Montalvo	16	16	0%	16	0%	16	0%	16	0%	16	0%	16	0%	16	0%	16	0%	16	0%	16
Jessica "Jess" Chong	17	17	0%	17	0%	17	0%	17	0%	17	0%	17	0%	17	0%	17	0%	17	0%	17
David Jelinsky	18	18	0%	18	0%	18	0%	18	0%	18	0%	18	0%	18	0%	18	0%	18	0%	18

ML Model Improvement Opportunities

v3 roadmap plan with areas for potential improvement

1. GridSearchCV hyperparameter tuning (Random Forest params, α, etc.)
2. Week-to-week predictions, as opposed to entire season
   - Pro: Better at handling varying strategies in early vs. mid vs. late game
   - Pro: Better approach to identify "goats" (i.e., less "threatening" players)
   - Con: Requires ~15x more data
3. Explore other features (ex: highest education level and/or profession)
4. Explore other ML model types

Supplemental Slides

Automated LLM-Analysis: Social & Strategy Scores

Social Score Criteria: Assesses the contestant's interpersonal skills, likability, and ability to navigate and influence social dynamics.

- Alliance Formation and Maintenance: The ability to create and maintain alliances that further the contestant's game.
- Social Integration: Effectiveness in becoming a key member of the group, avoiding social isolation.
- Jury Management: Skill in managing relationships with eventual jury members, crucial for securing votes in the final.
- Conflict Resolution: Competence in resolving disputes in a way that does not jeopardize their standing in the game.

Strategy Score Criteria: Assesses the contestant's game planning, tactical moves, and adaptability to changing dynamics.

- Strategic Planning: The ability to devise and implement plans that enhance their position in the game.
- Adaptability: Quick adjustment to new developments and ability to pivot strategies as the game evolves.
- Game-Changing Moves: Successfully executing moves that significantly alter the course of the game, including blindsides and effective use of immunity idols.

ML Model Feature Importances & Continuous Probability Transformation Functions

v2 Random Forest model feature importances

- num_idols_possessed: 0.077
- social_score: 0.090
- strategy_score: 0.239
- challenge_wins_per_day_lasted: 0.285
- confessionals_per_epi: 0.308

Exponential Decay:

$P(win) = e^{-\alpha \times (rank-1)}$

Where:

- $P(win)$ is the probability of winning.
- α is a parameter that controls the rate of decay, with higher values making the curve steeper.
- $rank$ is the contestant's final ranking.

Linear:

$P(win) = \frac{N-rank+1}{N}$

Where:

- $P(win)$ is the probability of winning.
- N is the total number of contestants in the season.
- $rank$ is the final ranking of the contestant.

Machine Learning Model Performances