On the Move:

A Sentiment Analysis on Professional Golfers' press conferences

Introduction to Text Mining and Natural Language Processing
Final Project

Luis Francisco Alvarez, Luke Atazona, Mikel Gallo

April 8, 2024

Abstract

We examine the sentiment from professional golfers' comments throughout the LIV and PGA dispute. Employing transcripts from press conferences, we construct sentiment features for professional golfers. We explore whether the sentiment in statements is a good indicator when constructing a prediction for players switching leagues.

Contents

1 Introduction

1.1 Context

The game of Golf influence, popularity and reach in the sport community has grown significantly over the years. In the USA for instance, a 2024 report by the National Golf Foundation (NGF) showed that there are currently about 27 million golfers and an estimated revenue from clubs and balls sales of close to USD 3 billion USD. Golf total reach is estimated at about 107 million people, more than one-third of the U.S. population The NGF has projected significant growth in the sport with much optimism and new activities, making golf an attractive labor market moving forward.
The Professional Golfers Association (PGA) was founded in 1901. The PGA is the oldest, most dominant and recognised professional golf association in the world, with a current global membership of over 28 thousand professional players worldwide. Its PGA Tour is the world's foremost professional golf league and home to the competition between virtually all the greatest golfers ever, including Jack Nicklaus, Tiger Woods and Arnold Palmer. In 2021 however, LIV Golf, backed and funded by the Saudi Public Investment Fund, emerged and disrupted the sport. LIV Golf lured several top PGA Tour players such as Phil Mickelson, Brooks Koepka, among others, with sidereal payouts. LIV also promised modernisation of the sport – music at events, looser dress codes, team competitions and three (as opposed to four) days lasting tournaments. In contrast to the PGA, LIV indicated it will not cut golfers with the worst scores after two rounds. This led to bitter relationships with golfers that joined LIV threatened severe sanctions and resignation from the PGA and DP World Tours. A series of legal disputes, litigations and tussles unfolded since this incursion in the professional golfing world. The Saudi government was accused of using LIV Golf to sway attention away from its human rights abuse and back-lashing players who joined the new league as money-chasers and unethical. Supporters of LIV argued however, that the PGA Tour has over the years deployed tacit tactics to cement its monopoly in the sport, seizing most of the revenue and content rights to detriment of the player's interests.

1.2 PGA and LIV Dispute

At its core, the conflict revolves around the PGA Tour's attempt to retain its players in the face of lucrative offers from LIV Golf, which is backed by the Saudi Public Investment Fund. This controversy has sparked a significant debate over loyalty, financial incentives, and the future direction of professional golf. Player's public comments offer a window into their evolving thoughts and loyalties. Over time, these comments have shifted from cautious to polarized stances, reflecting the broader tensions within the sport. Analyzing these sentiments could provide insights into the factors influencing players' decisions to stay with the PGA Tour or make the leap to LIV Golf, highlighting the complex interplay between financial considerations, personal values, and professional allegiance in this ongoing dispute. Professional golfers are among the richest athletes and stakeholders in sports, wielding special influence within the golfing community. Indeed, professional golfers' public opinions hold huge influence not only in the sport but in many other areas. For instance, Derenger

(2018) examined the endorsement effects of Tiger Woods, and found that Woods' endorsement led to over 28 thousand more *Titleist* club sales increasing revenues by USD 9.2 million (also see Elberse & Verleun, 2012). Golfers are also key influencers in social media. For instance, according to the 2024 Golf Influencers list, Rory McIlroy and and Tiger Woods have about 3 million followers each while Justin Thomas has 2 million on Instagram alone. Woods also has 6.6 million twitter followers and 3.2M on Facebook. These are suggestive that sentiments of professional golfers hold grand influence in the golfing world and this particular subject; PGA vs LIV.

1.3 Analytical Framework

We analyse the sentiment of professional golfers from the Tours in dispute, using a corpora capturing golfers public statements. In particular, we shall classify golfers' sentiments, identifying differences between golfers who switched Tours and those who did not, and how their commentary mutated through time. Their opinions, in general, tended to be very polarised in the beginning of the dispute, but gradually converged towards a more nuanced and balanced stand.

The rest of the paper is structured as follows. Section two provides a brief literature review. Section three discuss our data collections process, and exploratory data analysis. Section five presents our empirical strategy – model, and section six concludes with final results and thoughts, plus suggested areas for further research.

2 Related Literature

The use of sentiment analysis to extract, quantify and explore sentiments in texts is not new (see. e.g., Coase 1960; Markoff, et al., 1974; Markoff, 1982). For instance, Coase (1960) used legal texts to analyse issues of externalities. Recent advances in data, the availability of large corpora, improved methodologies, and NLP models such as Transformers have however revived and spurred interest in algorithmic text analysis (Gentzkow et al., 2017; Ash1 and Hansen, 2023). These have been reinforced by recent advances on "narrative economics", which has emphasized the role of narratives – stories and texts of individual and collective sentiments – in economics and the social sciences (Shiller, 2017; 2019). Shiller (2017) argue that sentiments reflected in popular stories can greatly enhance our predictive and explanatory power of major economic and social events. In their study, Aramaki et al., (2011) extracted influenza sentiments from tweets to detect influenza epidemics, finding high correlation (0.97) between tweets that mentioned influenzas especially at early stage of epidemics. Doh et al., (2021) examined the tone of FOMC statements finding it to correlate with medium-term policy expectations while Zahner (2021) analyzed sentiments expressed in public speeches of the ECB over 2002 -2020. He finds inflation rate of "above, but close to 2%". In a recent study on the effects of media coverage on economic events, Besley e al., (2024) developed a model to examine how events are driven by media coverage. They fitted their model on a large corpus of news data on five countries reporting on violent events in the destinations and concluded that media coverage can triple economic impact of an event".

3 Data

3.1 Data Source

The Data we will be working with is a recompilation of press conference transcripts, where players of our interest answer previously asked questions, after rounds of play. All the resources are found at ASAP Sports. The period we analyse is from the 2020 Augusta Masters' Championship held in early November of that year up until the present day (March 7th 2024). It is relevant to note that at that point in time, LIV was not a reality and the PGA Tour was the only top league in the golfing world. The players chosen are both from LIV and the PGA Tour. For those that actually switched tours, their departure date is also relevant for our analysis. They were selected on the basis of their relevance, in terms of their number of titles, golfing resume and how vocal they were during the confrontation once the LIV plan unfolded. Most of the questions are asked to players who performed well during that tournament; this could be a source of bias that is worth to have in mind. Many of the question-answers are directed towards the actual game such as state of the golf field or the weather, the tournament or shots in particular. A much more robust analysis can be carried on by expanding the pool of scraped players, or identifying other sources from where players could be voicing their opinions (Ex: Social Media).

3.2 Scraping

As mentioned above, we have decided to scrape the press conference transcripts for all selected players. This will involve retrieving the transcripts in which they participated, parsing the data to extract the date, questions asked, the respondent, and the player's actual response. The final dataset, is created upon 8952 press conferences, from our original 16 players. We used *Beautiful soup* to search through them and were able to parse 19352 question - answer pairs with the player responding.

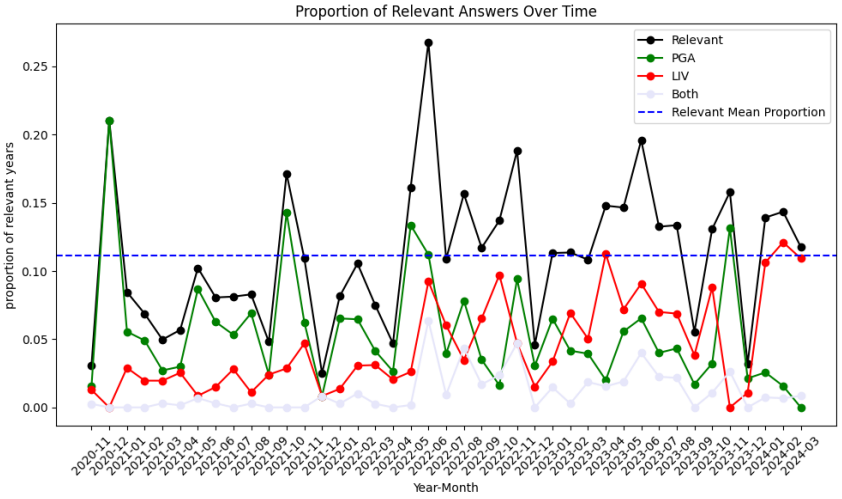
3.3 Exploratory Data Analysis

We first identify which answers we will consider relevant, out of the 19352, we keep 2209 for our by player analysis. Although this seems to be a low proportion of the total dataset, we have to consider that the original data is from post-play interviews. We consider that our more than 2200 question-answers pairs are a good starting point for the sentiment analysis. The proportion of answers that are identified as relevant out of the whole set of questions for that point in time. We also dissect from our relevant answers, which ones were identified by each league.

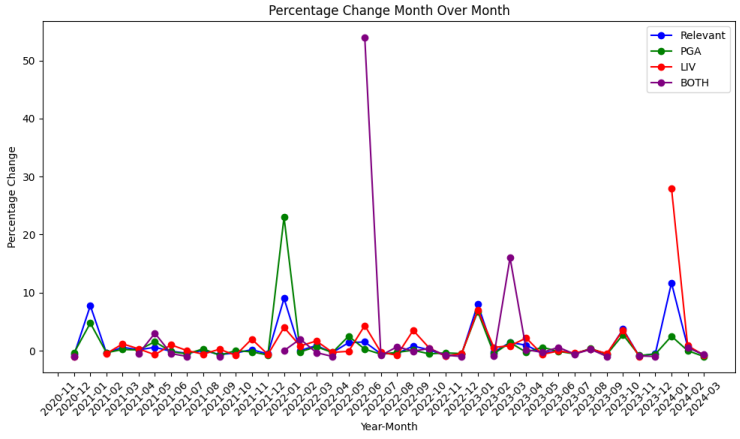
Spikes can be linked to real world events in the golfing sphere. For example, the highest point in our series coincides with the launch of the first LIV golf event in May 2022. Just a couple months later we still see high number of relevant observations, they can be associated with the announcement of departure from the PGA Tour of an important pool of players such as Brooks Koepka, Bryson Dechambeau and Joaquin Niemann. Another high value, like April 2023 is most likely due to the 2023 Augusta National Masters tournament, the

first one with both LIV and PGA players in the field since the 2022 fracture.

In this graph we differentiate by topic. For the first half of our timeline, mentions of the PGA tour (represented by the green series) dominate. However, as time progresses, the LIV league gains prominence, with its mentions (shown in red) surpassing those of the PGA, specially in the final quarter of our time period.
Finally, it is important to note that generally, players who performed well are interviewed, then another analysis could be done to other players. Perhaps "worse" players have different results.

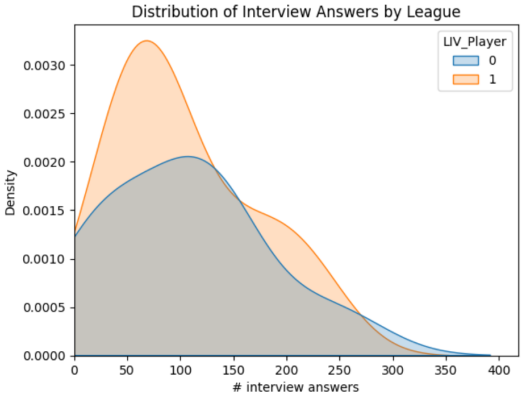


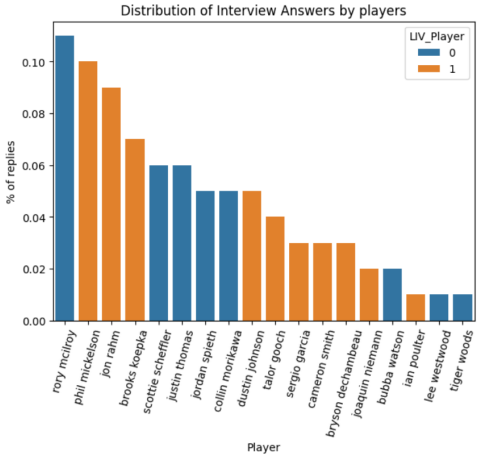
In this second plot, we observe the first difference of the time series, it is clearer to identify big unprecedented changes in proportions. Towards the end of 2023, we see a big increase in the LIV mentions; this aligns with the agreement and announcement of Jon Rahm, arguably the best player in the world, to also join the younger tour and providing a new big blow to the PGA Tour roster of players.



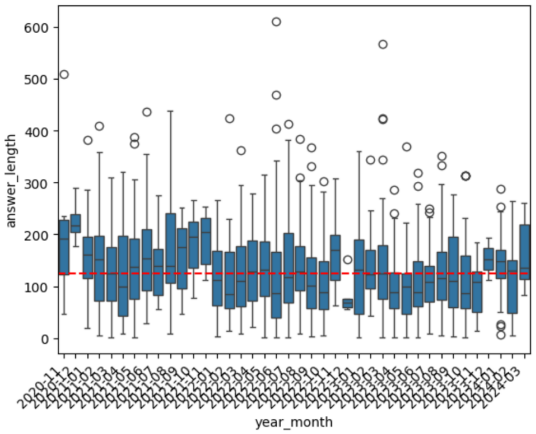
We will also delve into the distribution of answers (and their length) by players to see if

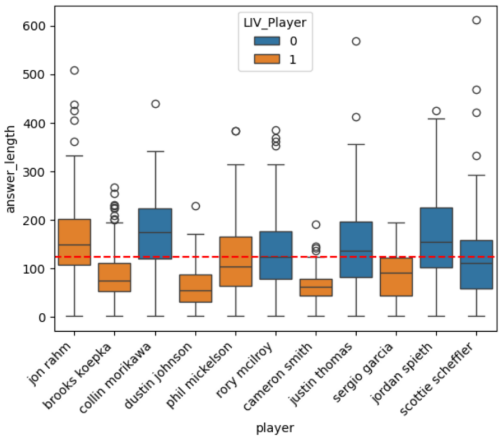
there are any relevant differences to consider between the type of players.





This visual suggests that, on average, LIV players tend to have fewer interview answers, which aligns with the information that they participate in fewer tournaments, leading to fewer media interactions. Furthermore, the LIV player field is substantially shorter, which contributes to the same players being questioned more frequently. As a result, the density of their responses is higher at the lower end of the scale, indicating a more consistent, but overall reduced, presence in press conferences compared to their counterparts.

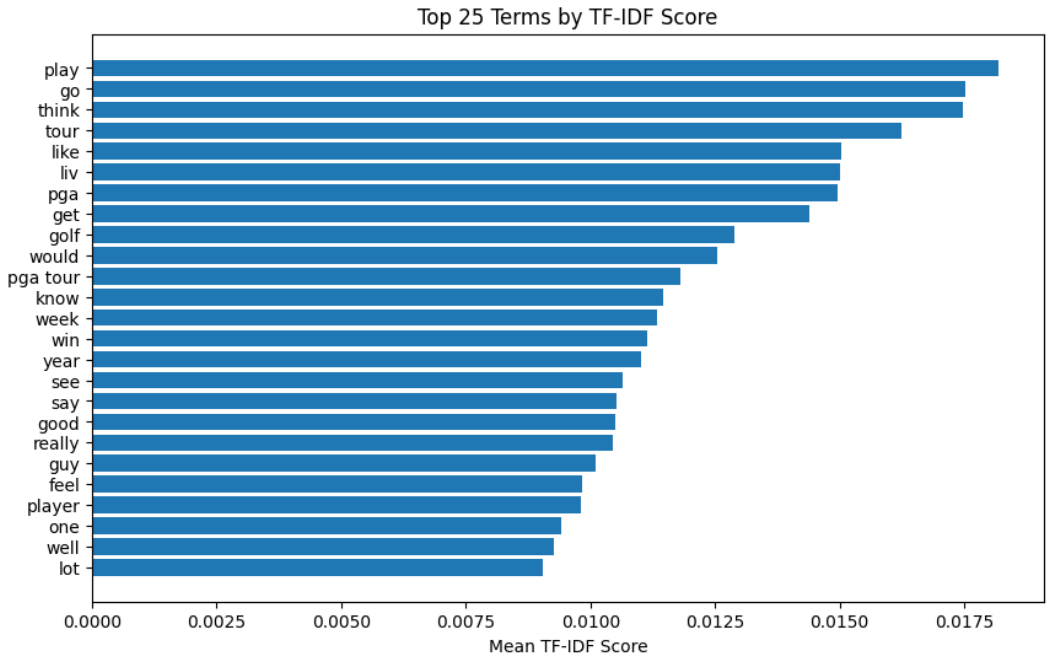




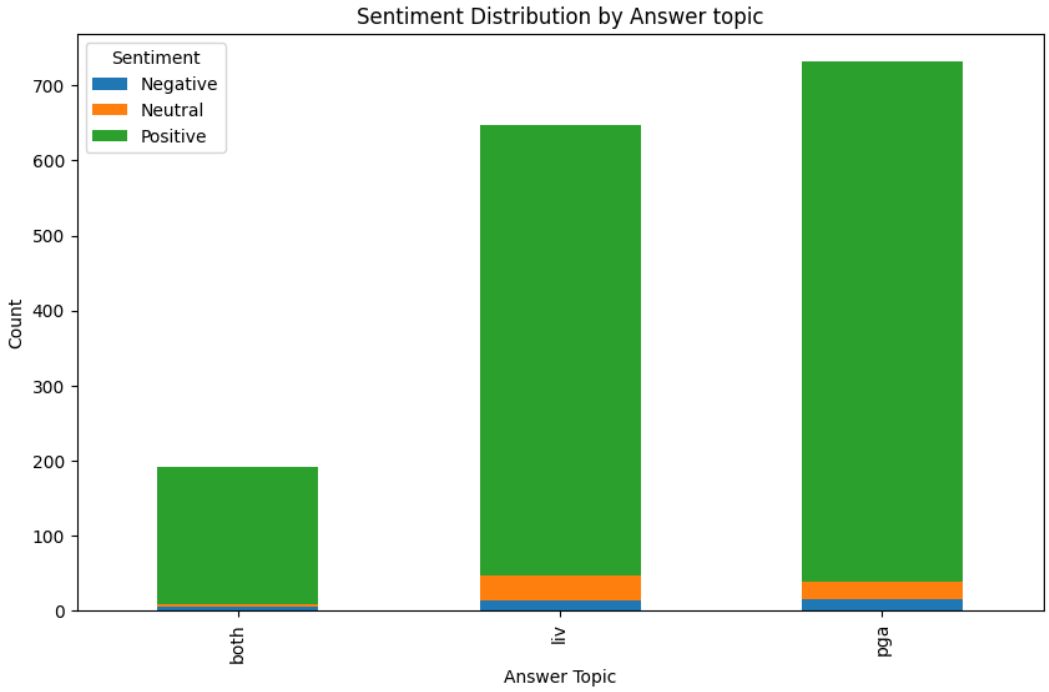
Once again, PGA players have more observations. Note that Jon Rahm is shown as a LIV player but his switch was during January 2024, so most of his press conferences were during his PGA Tour membership. The data sample is however, fairly balanced. The average length of players' response does not seem to be indicative of any difference a priori.

4 Sentiment Analysis

Before conducting the sentiment analysis, we prepared the interview responses by applying a series of preprocessing steps. This involved converting all text to lowercase, eliminating stopwords, and lemmatizing the words. This preparation was essential for aligning the words with a sentiment model, ensuring capturing the sentiment expressed in the interviews. Additionally, we added two layers on top: CountVectorizer to compute the frequency of our terms and then applied a tf-idf transformation to highlight the most significant terms.

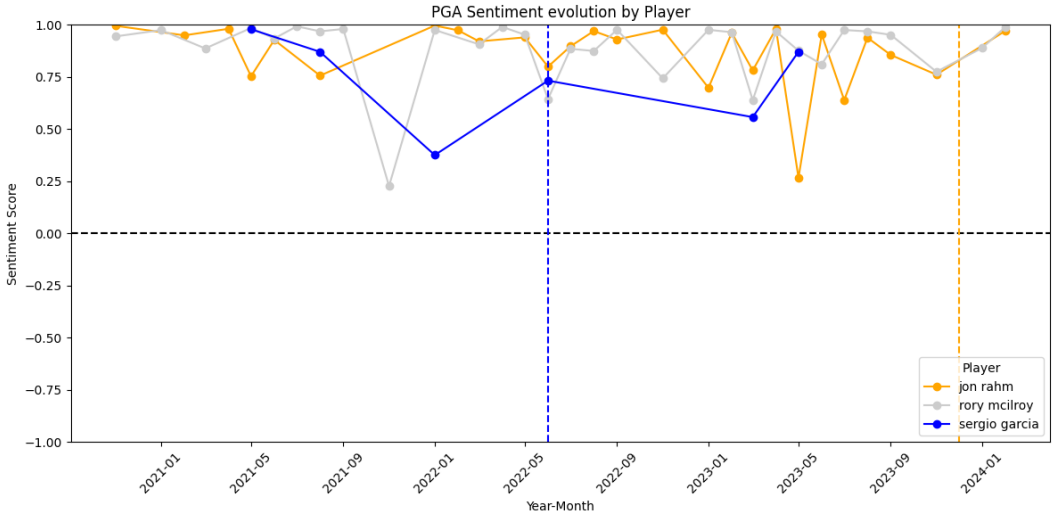


In our sentiment analysis, we used VADER (Valence Aware Dictionary for Sentiment Reasoning), a model known for its sensitivity to both the polarity (positive and negative) and intensity (strength) of text's emotions. Originally made for social media sentiment analysis, we found it applicable to our needs, particularly in interpreting transcripts from live interviews and press conferences. Furthermore, the scores range from -1 to 1, where -1 signifies a very negative sentiment, 0 stands for neutral, and 1 indicates a highly positive sentiment.

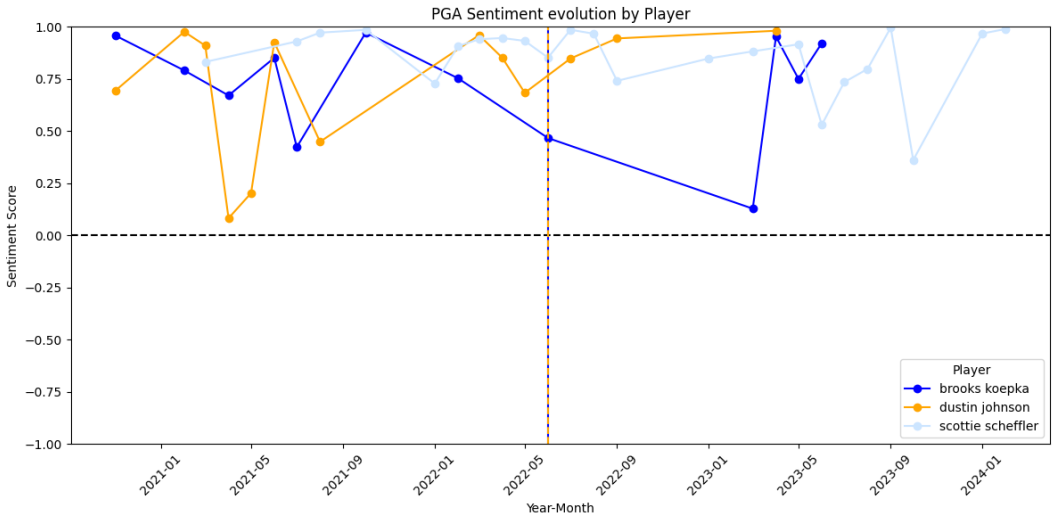


The sentiment conveyed in the interviews tends to lean heavily towards positivity in all three topics. This outcome wasn't exactly surprising given the circumstances surrounding player interviews. Most of the players interviewed were finalists and winners of the tournaments, which naturally leads to speeches filled with positivity and encouragement. Another noteworthy aspect to consider might be that players may not want to openly criticize their current affiliations, which could potentially have negative consequences.

In this section, we took a deep dive into how players' sentiments toward the PGA Tour have evolved since the official announcement of the LIV league back in 2020, all the way up to early 2024. Our goal is to uncover any signs in player sentiment that could hint at a potential shift towards the LIV competition. Now, we're about to showcase a couple of player groups (a mix of both PGA and LIV players) to shed light on some patterns that we've spotted among those who switched leagues compared to those who stayed.



In Cohort 1, we closely observed the sentiment journey of both Sergio Garcia and Jon Rahm, who made the switch to LIV in 2022/23. Interestingly, Sergio's sentiment showed a steady decline in the three months leading up to his transition, suggesting that he might have been signaling his intentions and adjusting his tone towards the PGA prior to his departure. On the other hand, Rahm's case presents a different scenario, his sentiment remained consistently positive until the very end, when he suddenly moved to LIV, especially noteworthy considering the extraordinary circumstances surrounding his decision, including a hefty welcome package of 800 million. Nonetheless, he kept a consistent pattern similar to Rory McIlroy, who opted to stay in the PGA.

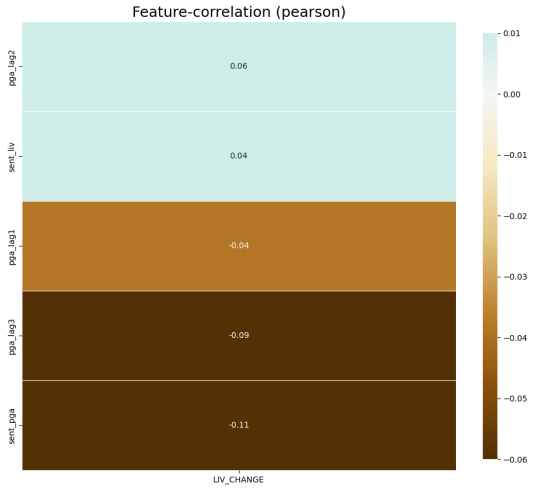


In Cohort 2, we notice a familiar pattern with Brooks Koepka, exhibiting a gradual decline in sentiment leading up to his shift to LIV. However, upon examining Dustin Johnson's trend, there are no evident signs of discontent or shifts in tone prior to his change of affiliation.

This particular case could either be an outlier similar to what we observed with John Rahm, or it might suggest that there's no clear-cut pattern linking lower sentiment with a higher probability of switching leagues.
It's crucial to note that our research was conducted with a very limited pool of players and interviews. This limitation could result in increased variability and, consequently, reduce the reliability of our findings.

5 Modelling

In this final section of the analysis, we look into sentiment as a predictor for the switch by players from tours. For this, we not only consider LIV and PGA sentiment, but also the PGA lags since it holds more information. Since we are working with few players in general, even less those who made the change, we do a re balancing strategy, *Random Over Sampling* so we can obtain a better set of predictions. This is the initial correlation matrix, where the PGA sentiment hold the most negative correlation and LIV sentiment appears to be uncorrelated to our flag variable:



We train a Logistic regression for the classification exercise.

Logit Regression Results

Dep. Variable:	LIV_CHANGE	No. Observations:			317	
Model:	Logit	Df Residuals:			311	
Method:	MLE	Df Model:			5	
Date:	Mon, 18 Mar 2024	Pseudo R-squ.:			0.2227	
Time:	22:27:14	Log-Likelihood:			−122.86	
converged:	True	LL-Null:			−158.07	
Covariance Type:	nonrobust	LLR p-value:			8.378e-14	
	coef	std err	z	P>|z|	[0.025	0.975]
const	−1.8852	0.205	−9.189	0.000	−2.287	−1.483
x1	−0.8254	0.166	−4.977	0.000	−1.150	−0.500
x2	−0.2431	0.145	−1.673	0.094	−0.528	0.042
x3	0.8856	0.308	2.877	0.004	0.282	1.489
x4	−0.5347	0.151	−3.546	0.000	−0.830	−0.239
x5	0.6383	0.239	2.670	0.008	0.170	1.107

The predictions where done in-sample. However, we proposed a way moving forward. By fitting and ARIMA process to the series and generating new observations, we could test

the model's predictive power in identifying when a player is close to move from the tour. Naturally this observations would be synthetic ones.

6 Results and Interpretations

This research paper marks our first step in attempting to forecast the future of the PGA TOUR, the LIV TOUR, and the players navigating this challenging conflict.
Reflecting on the limitations of our model we conclude that due to time and resources limitations, we restricted our scraping to few players, hence limiting the power of out experiment. A way the model could be extended is by gathering data from a bigger pool of players and using a subset of them as a validation/test dataset. We opted against it due to the fact that we were working with very scarce observations. Ideally, the training of the model should only be carried on observations before a player switches tours and tested on independent ones. Once again, we decided to not subset our data any further.

References

- Ash, E., & Hansen, S. (2023). *Text Algorithms in Economics.* *Annual Review of Economics, 15*, 659–88.
- Aramaki, E., Maskawa, S., & Morita, M. (2011). *Twitter Catches the Flu: Detecting Influenza Epidemics using Twitter.*
- Besley, T., Fetzer, T., & Mueller, H. (2024). *How big is the media multiplier? Evidence from dyadic news data.* The Review of Economics and Statistics. https://doi.org/10.1162/rest_a_01415
- Coase, R. H. (1960). *The Problem of Social Cost.* The Journal of Law & Economics, 3, 1–44. http://www.jstor.org/stable/724810
- Chung, K. Y. C., Derdenger, T. P., & Srinivasan, K. (2013). *Economic Value of Celebrity Endorsements: Tiger Woods' Impact on Sales of Nike Golf Balls.* Marketing Science, 32 (2), 271-293. https://doi.org/10.1287/mksc.1120.076
- Derdenger, T. P. (2018). Examining the impact of celebrity endorsements across consumer segments: An empirical study of Tiger Woods' endorsement effect on golf equipment. *Marketing Letters, 29*(2), 123-136.
- DiMarco, J. (2023). PGA Golf players and LIV Golf: Communication and Image Repair Strategies. In *Sage Business Cases*. SAGE Publications, Ltd. https://doi.org/10.4135/9781529620207
- Doh, T., Kim, S., & Yan, S.-K. (2021). How You Say It Matters: Text Analysis of FOMC Statements Using Natural Language Processing. *Economic Review, 2021.* Federal Reserve of Kansas City.
- Elberse, A., & Verleun, J. (2012). The economic value of celebrity endorsements. *Journal of Advertising Research, 52*(2), 149–165.
- Gentzkow, M., Kelly, B. T., & Taddy, M. (2017). Text as data. *NBER Working Paper No. 23276.*
- Gulf Digest. (2014). https://www.golfdigest.com/
- Hajek, P., & Henriques, R. (2024). Predicting M&A targets using news sentiment and topic detection. *Technological Forecasting and Social Change, 201*, 123270. https://doi.org/10.1016/j.techfore.2024.123270.
- National Golf Foundation (various). Golf Participation Report. https://www.ngf.org/golf-participation-update-bigger-younger-and-cooler/
- Markoff, J., Shapiro, G., & Weitman, S. (1974). Toward the integration of content analysis and general methodology. In D. R. Heise (Ed.), *Sociological Methodology, 1975* (pp. 1–58). Jossey-Bass.

- Markoff, J. (1982). Suggestions for the Measurement of Consensus. *American Sociological Review, 47*(2), 290–298. https://doi.org/10.2307/2094970
- Shiller, R. J. (2019). Narrative Economics: How Stories Go Viral and Drive Major Economic Events.
- Zahner, J. (2021). Above, but close to two percent. Evidence on the ECB's inflation target using text mining. School of Business and Economics, Institutional Economics Research Group, Philipps-Universität Marburg, Germany.
- Most Followed Golfer on Social Media. https://essential.golf/most-followed-golfers-on-social-media/