

Final Report - Work in Progress

- Research Hypothesis / Questions:
 - Is Formula 1 fandom Toxic?
 - Are there specific groups that show more toxic behaviour then others?
 - Is the toxicity a “self-made” problem of Formula 1?
- APIs: Youtube
 - (Not reddit as post are often off topic especially during the off season, that we are currently in)
- Methods:
 - TBD
 - Dictionary
 - * Formula 1 specific words that are toxic
 - * racism / ethnic slurs -> (“List of Ethnic Slurs,” 2023)
 - * toxicity -> (Orthrus-Lexicon, 2022)
 - * hate speech -> (Vegt et al., 2021)
 - * insults -> (Vegt et al., 2021)
 - Transformer classifier
 - * sentiment -> cardiffnlp/twitter-roberta-base-sentiment-latest (Barbieri et al., 2020)
 - * racism -> jaumefib/datathon-against-racism (Aluru et al., 2020)
 - * hate speech -> Hate-speech-CNERG/dehatebert-mono-english (*Jaumefib/Datathon-Against-Racism* · *Hugging Face*, n.d.)
 - statistical analysis
 - * group toxic behavior by drivers and teams
 - * group by topics
 - topic modelling?
- Contents:
 - Introduction
 - * What is Formula 1
 - * Why do we need to analyze this
 - * introduce the three research questions / hypothesis
 - Fundamentals
 - * What is fandom
 - * Defining toxic fan behavior
 - * Youtube API
 - * Maybe explaining the used methods?
 - Concept
 - * What will be done
 - * How will i be doing it
 - Creating the Dataset
 - * Explain Dataset creation
 - Applying Method 1
 - Applying Method 2
 - Results

Analysing Toxicity in Formula 1 Fandom - Computational Analysis of Communications Final

Author: Leon Knorr

Matr-Nr: 1902854

Introduction

Formula 1 is the highest class of international racing for open-wheel single-seater formula racing cars and is generally considered the most competitive, fastest and hardest class of motor racing. Since it's first season in 1950, Formula 1 is visiting a diverse list of many different countries, where the best drivers in the world are racing against each other in teams of two drivers to determine the best driver and the best team on the Formula 1 grid (Lowrey, 2019). These events are visited by thousands of Fans, with millions more following them on television and social media. With the 2021 season being one of the closest and most entertaining seasons in the history of Formula 1, where Red Bulls Max Verstappen beat Mercedes driver Lewis Hamilton in the grand finale of the season under controversial circumstances after a full season of controversy, drama and intense on track battles and with the release of Netflix Drive To Survive, Formula 1s popularity is growing rapidly. But, reports of Toxic and abusive Fan behavior at events and in comment sections on social media are accumulating, and casts an ugly shadow over Formula 1s latest successes (Woodhouse, 2022). As the reports over toxic and abusive fan behaviours in social media and at live events are rising, Formula 1 as well as Fans and drivers are taking a stand against toxicity in the Formula 1 community. However, an independent and scientific analysis of this topic is missing and therefore the accusations are sort of hanging in the air without a solid scientific foundation. Therefore, in order to tackle this problem research into the toxicity of Formula 1 fandom is a necessity to gain valuable insights into understanding the problem, where it originates from and to build a foundation for future measures to make attending Formula 1 events as well as the media around it a safer and more enjoyable experience. To take the first step into this direction, this thesis will analyse Youtube comments of the Formula 1 channel in order to determine:

- If the Formula 1 fandom is toxic
- Are there specific groups that are more toxic than others?
- Is the toxicity a “self-made” problem of Formula 1 and where is the toxicity originating from?

Fundamentals

Concept

The Dataset

Dictionary Analysis

Transformer Classifiers

Results

Bibliography

““

- Aluru, S. S., Mathew, B., Saha, P., & Mukherjee, A. (2020). *Deep Learning Models for Multilingual Hate Speech Detection*. arXiv. <https://doi.org/10.48550/arXiv.2004.06465>
- Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). *TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification*. arXiv. <http://arxiv.org/abs/2010.12421>
- Jaumefib/datathon-against-racism* · *Hugging Face*. (n.d.). Retrieved January 9, 2023, from <https://huggingface.co/jaumefib/datathon-against-racism>
- List of ethnic slurs. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_ethnic_slurs&oldid=1132041045
- Lowrey, T. (2019). *About F1 | Formula One World Championship Limited*. <https://corp.formula1.com/about-f1/>
- Orthrus-Lexicon. (2022). *Orthrus Toxic Dictionary implementation*. <https://github.com/Orthrus-Lexicon/Toxic>
- Vegt, I. van der, Mozes, M., Kleinberg, B., & Gill, P. (2021). The Grievance Dictionary: Understanding threatening language use. *Behavior Research Methods*, 53(5), 2105–2119. <https://doi.org/10.3758/s13428-021-01536-2>
- Woodhouse, J. (2022). “Scary” how toxic Formula 1 community became during 2021. In *PlanetF1*. <https://www.planetf1.com/news/toxic-formula-1-community-2021/>