Subreddit Classification and Sentiment Analysis using NLP:
AMD vs NVIDIA

» TABLE OF CONTENTS «

Introduction

01

Introduction

Background:

AMD Radeon GPUs are frequently seen as an inferior, **“budget”** option compared to nVidia GeForce GPUs. Aside from pricing, a common complaint is that **AMD GPU drivers are buggy and unreliable.**

Introduction

The AMD Strategic Planning (Consumer GPUs) team wants our team to find out whether the “**negative stigma**” associated with AMD stems from “**genuine issues**” or if it is simply a matter of “**perception**”.

02

Problem Statement

The perception of Radeon vs Geforce

Disclaimer

Means colourful words

PROBLEM STATEMENT

Problem:

It is unclear whether the **perception of AMD** products being more problematic is a **marketing** or an **engineering issue**. Some light can be shed on this through sentiment analysis.

Must be Engineering issue !! Engineering 🍍

Must be Marketing Issue! Cause marketing Department 🍍

PROBLEM STATEMENT

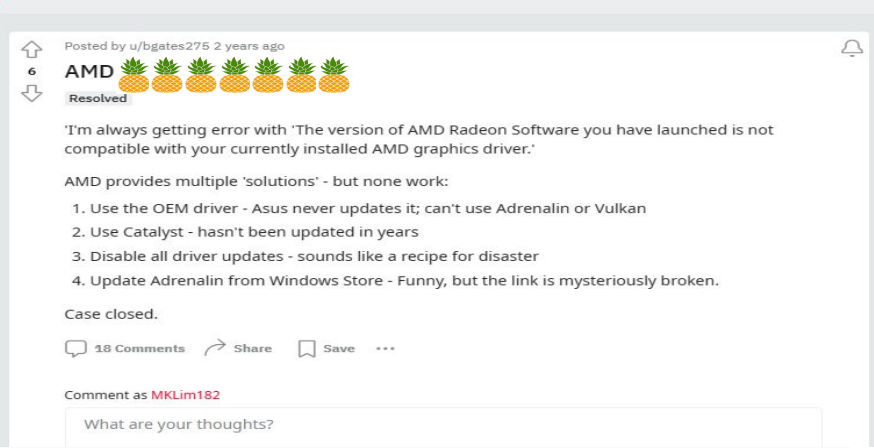
Theory:

1. Subreddits are moderated - negative posts without specific reference to a problem gets removed.

2. Example of negative post will get removed “AMD 🍍🍍🍍🍍🍍🍍🍍”

3. Does the above post mention anything specific to the problem?

PROBLEM STATEMENT



Posted by u/bgates275 2 years ago

6

AMD

Resolved

'I'm always getting error with 'The version of AMD Radeon Software you have launched is not compatible with your currently installed AMD graphics driver.'

AMD provides multiple 'solutions' - but none work:

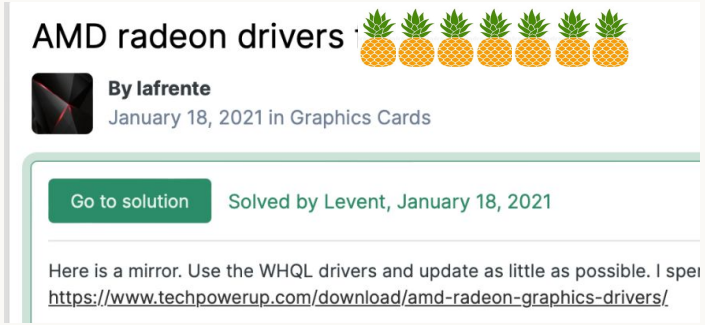
1. Use the OEM driver - Asus never updates it; can't use Adrenalin or Vulkan
2. Use Catalyst - hasn't been updated in years
3. Disable all driver updates - sounds like a recipe for disaster
4. Update Adrenalin from Windows Store - Funny, but the link is mysteriously broken.

Case closed.

18 Comments Share Save ...

Comment as MKLim182

What are your thoughts?



AMD radeon drivers

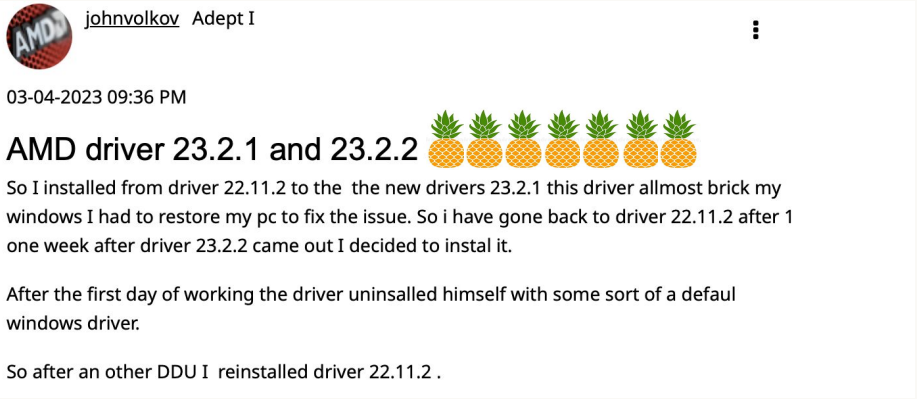
By Iafrente

January 18, 2021 in Graphics Cards

Go to solution Solved by Levent, January 18, 2021

Here is a mirror. Use the WHQL drivers and update as little as possible. I spe

https://www.techpowerup.com/download/amd-radeon-graphics-drivers/



johnvolkov Adept I

03-04-2023 09:36 PM

AMD driver 23.2.1 and 23.2.2

So I installed from driver 22.11.2 to the the new drivers 23.2.1 this driver allmost brick my windows I had to restore my pc to fix the issue. So i have gone back to driver 22.11.2 after 1 one week after driver 23.2.2 came out I decided to instal it.

After the first day of working the driver uninsalled himself with some sort of a defaul windows driver.

So after an other DDU I reinstalled driver 22.11.2 .

PROBLEM STATEMENT

Theory:

1. We would expect there only to be negative posts about actual problems.

2. If a higher proportion of posts on r/amd is negative than on r/nvidia, likely there is genuine issue with AMD QA.

3. If a similar proportion of posts on both subreddits are negative, then it is a perception issue.

PROBLEM STATEMENT

Scope:

1. Use text data posted on both subreddits to build a model to differentiate between text about them.

2. The idea is to use the model on other GPU discussion forums with no split between posts about AMD/nVidia to pick out posts about AMD to apply the same sentiment analysis tools we are applying to the Reddit dataset.

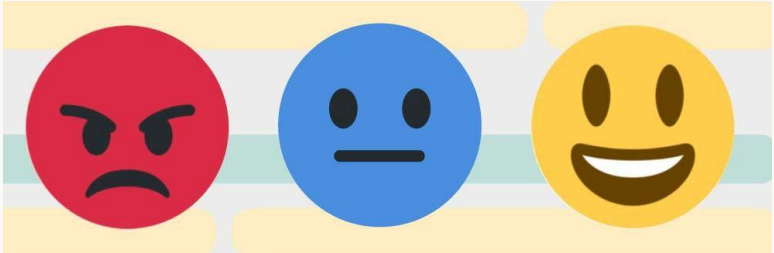
3. Use sentiment analysis tools to establish whether the perception is driven by nVidia's mindshare or by a genuine engineering problem.

PROBLEM STATEMENT

Success Metrics:

1. **Precision** - we want to ensure that the vast majority of the posts our model flags as “amd” are actually AMD-related.

2. **VADER Sentiment Analysis** results (positive or negative sentiment) to recommend whether to deploy more resources to marketing or engineering.

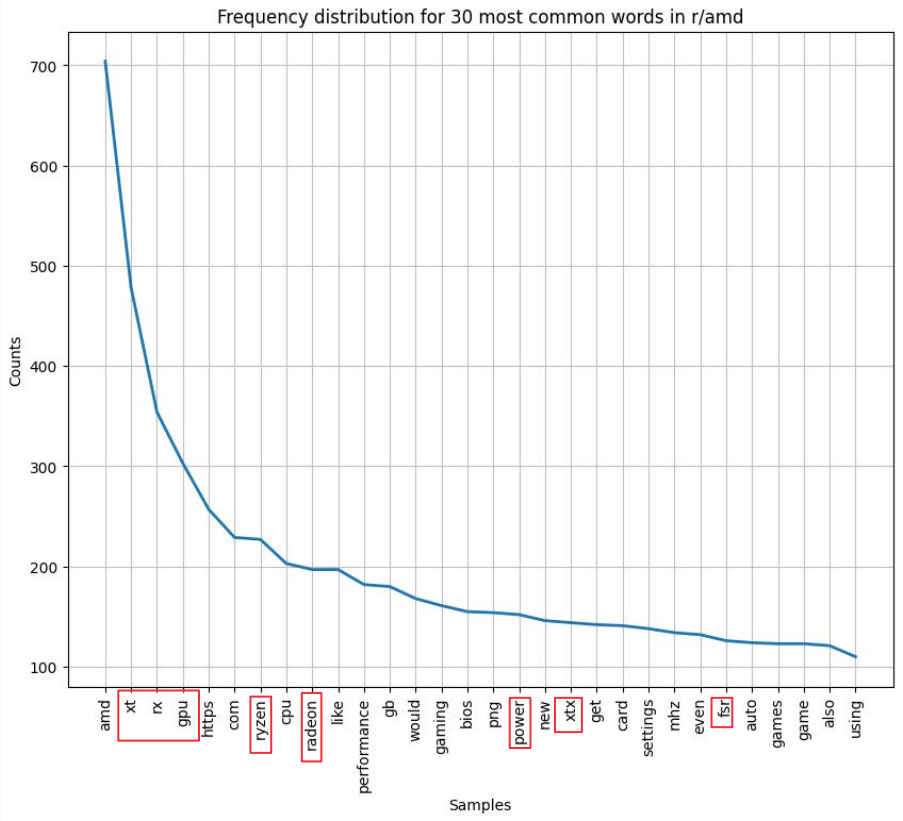


EDA

04

This dataset contains a large amount of subject-matter specific vocabulary.

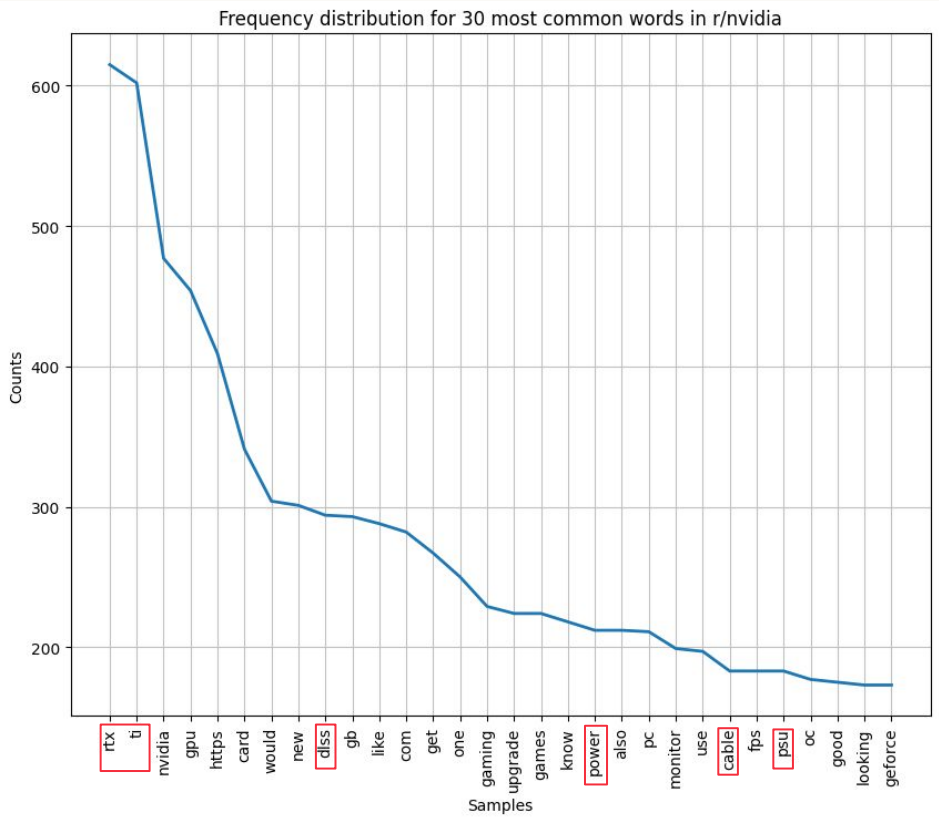
EDA - Frequency Distribution



RADEON GRAPHICS

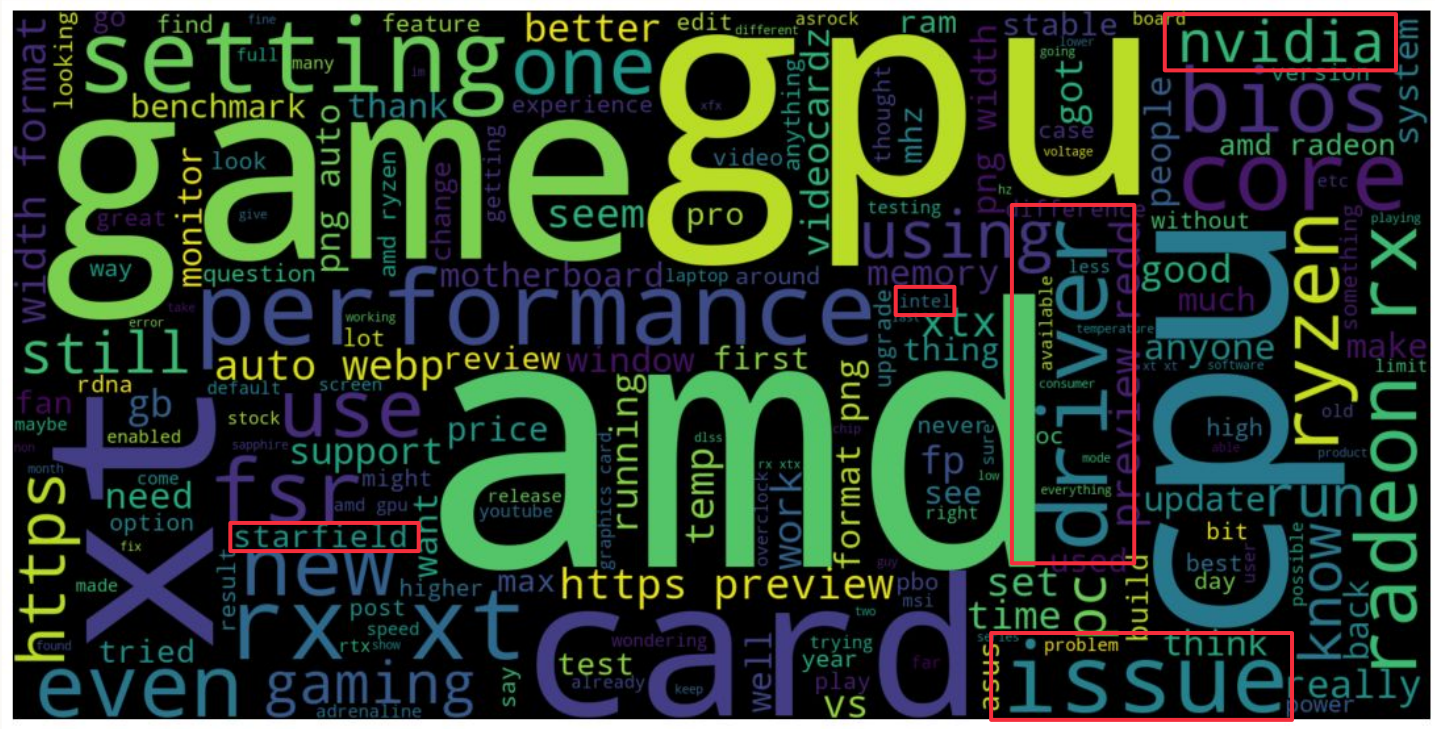
AMD

EDA - Frequency Distribution



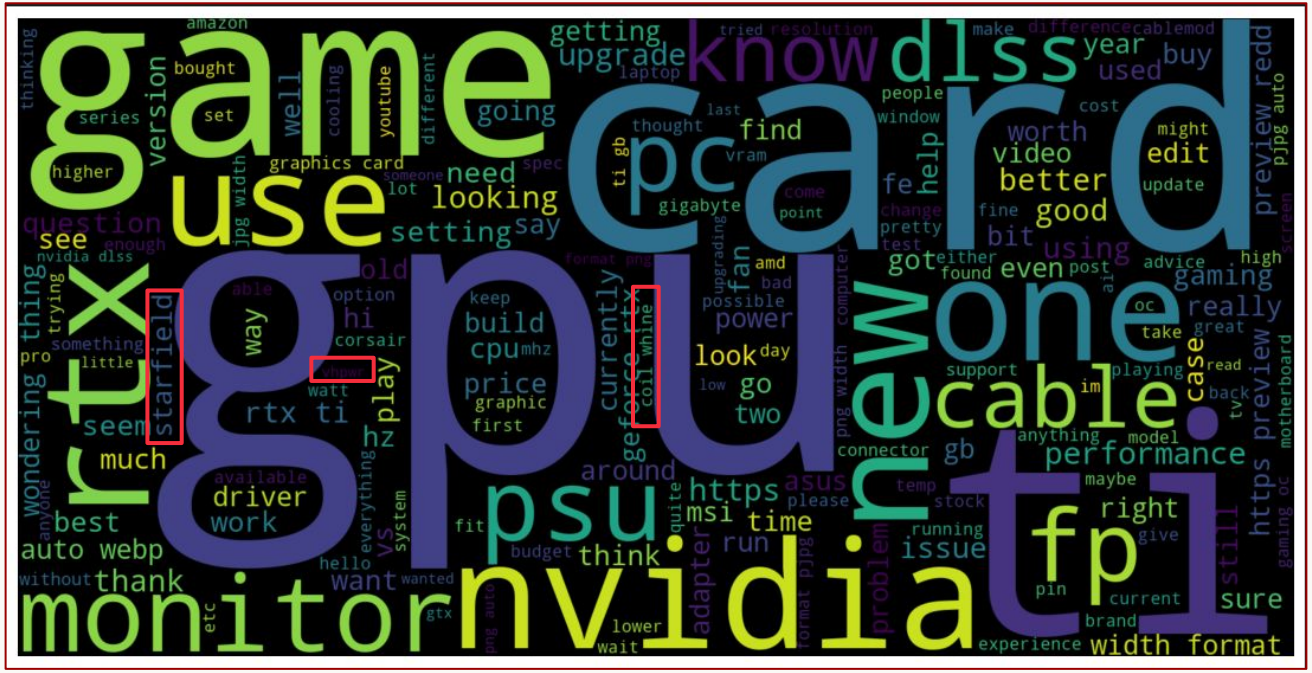
NVIDIA GEFORCE

EDA - Word Cloud



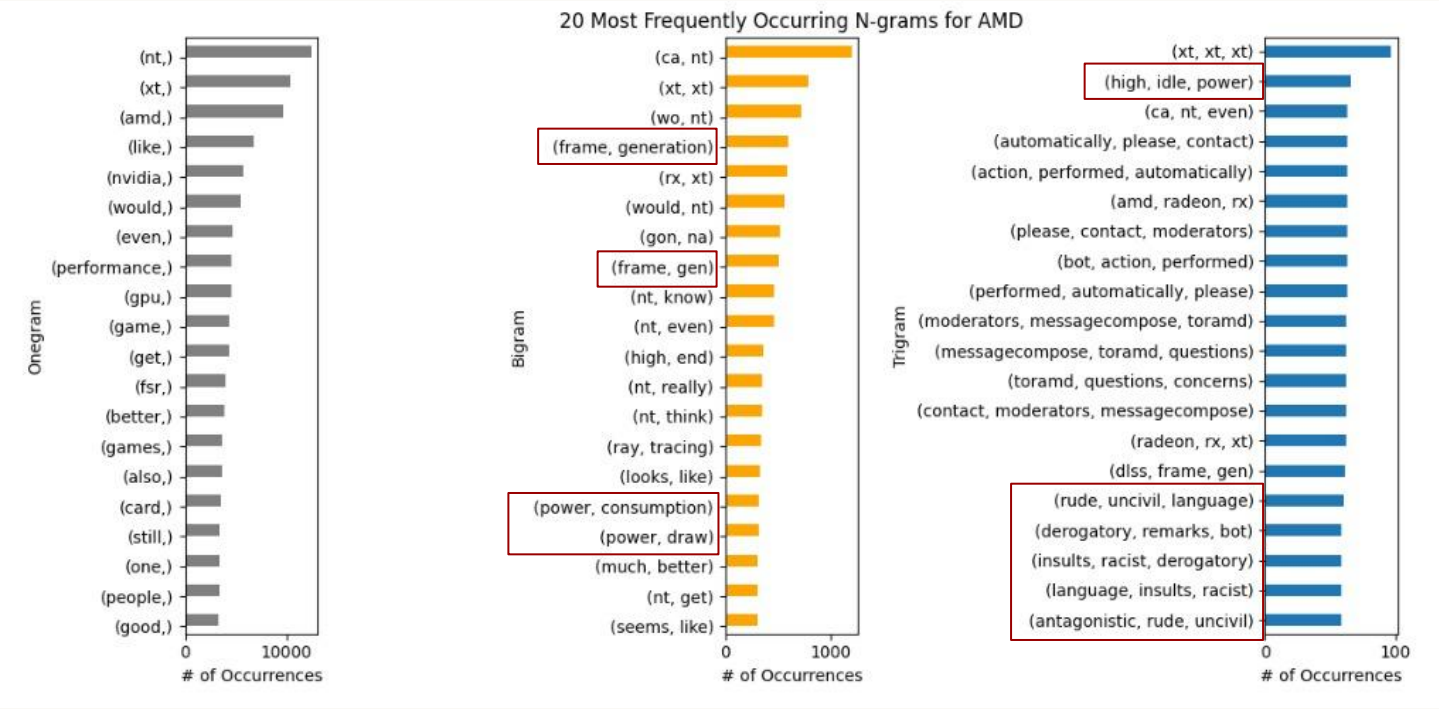
RADEON GRAPHICS AMD

EDA - Word Cloud

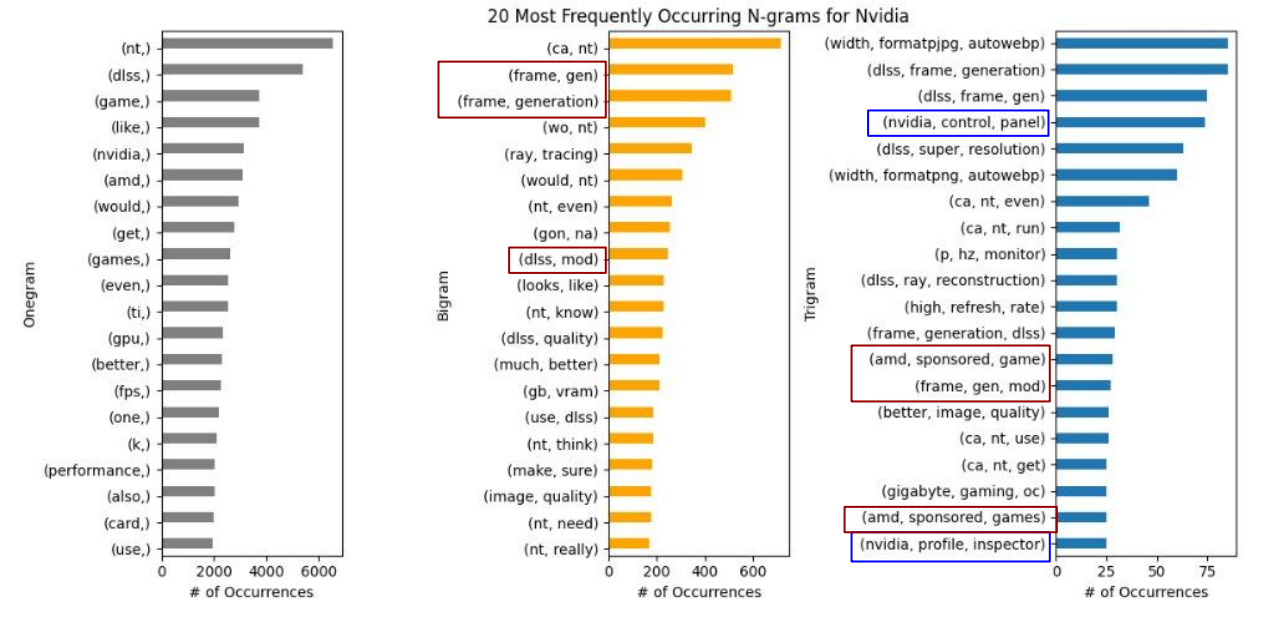


NVIDIA GEFORCE

EDA - N-Grams


20 Most Frequently Occurring N-grams for AMD
Onegram
(nt,)
(xt,)
(amd,)
(like,)
(nvidia,)
(would,)
(even,)
(performance,)
(gpu,)
(game,)
(get,)
(fsr,)
(better,)
(games,)
(also,)
(card,)
(still,)
(one,)
(people,)
(good,)
0
10000
# of Occurrences
Bigram
(ca, nt)
(xt, xt)
(wo, nt)
(frame, generation)
(rx, xt)
(would, nt)
(gon, na)
(frame, gen)
(nt, know)
(nt, even)
(high, end)
(nt, really)
(nt, think)
(ray, tracing)
(looks, like)
(power, consumption)
(power, draw)
(much, better)
(nt, get)
(seems, like)
0
1000
# of Occurrences
Trigram
(xt, xt, xt)
(high, idle, power)
(ca, nt, even)
(automatically, please, contact)
(action, performed, automatically)
(amd, radeon, rx)
(please, contact, moderators)
(bot, action, performed)
(performed, automatically, please)
(moderators, messagecompose, toramd)
(messagecompose, toramd, questions)
(toramd, questions, concerns)
(contact, moderators, messagecompose)
(radeon, rx, xt)
(dlss, frame, gen)
(rude, uncivil, language)
(derogatory, remarks, bot)
(insults, racist, derogatory)
(language, insults, racist)
(antagonistic, rude, uncivil)
0
100
# of Occurrences

EDA - N-Grams



NVIDIA GEFORCE

Modelling

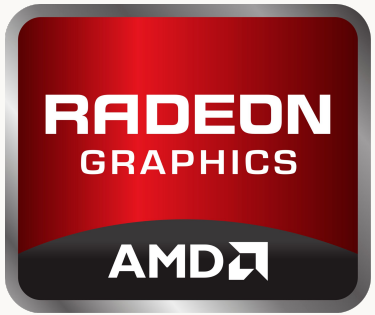
Powered by: 

04

Binary Classification:

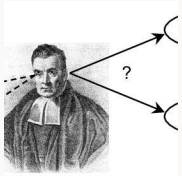
Does a post belong to r/AMD or r/Nvidia?

AMD



Nvidia

Types of Models for CLASSIFICATION tasks



Naive Bayes

Simple probabilistic classifier based on Bayes' theorem

Random Forest

Ensemble of decision trees for high accuracy

Adaboost

Boosts weak learners for improved performance

Neural Net

Deep learning model inspired by the human brain, for complex tasks

Types of Text VECTORISERS

Count Vectoriser

Represents texts using word frequency

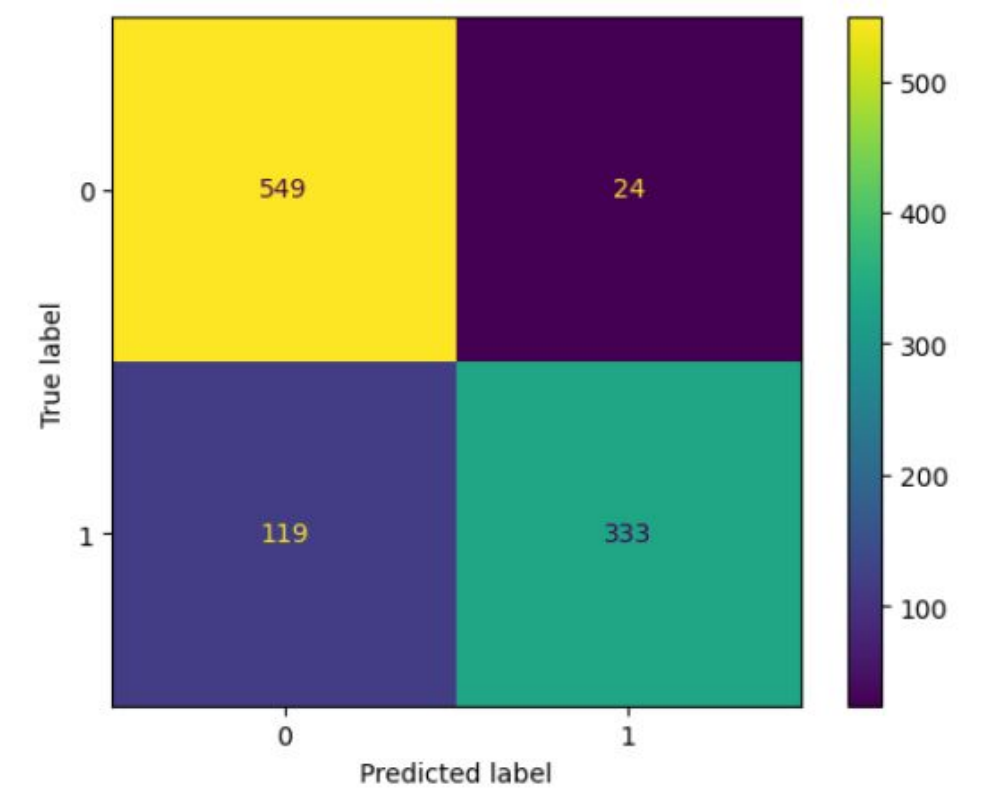
TF-IDF Vectoriser

Compares the number of times a word appears in a documents vs number of documents the word appears in

Random Forest with TFIDF Vectoriser was selected as the best model

Model	Train Score	Test Score	Precision	Recall	F1-Score	Fit time
Naive Bayes with cvec	0.990	0.944	0.909	0.820	0.862	18s
Naive Bayes with tvec	0.993	0.941	0.915	0.739	0.818	14s
Random Forest with tvec	0.999	0.937	0.932	0.736	0.823	43s
Adaboost with tvec	0.999	0.913	0.885	0.735	0.819	58s
Sequential Neural Net with tvec	N/A	0.865	0.931	0.752	0.832	5s (CUDA)

876 out of 1025 accurate predictions overall



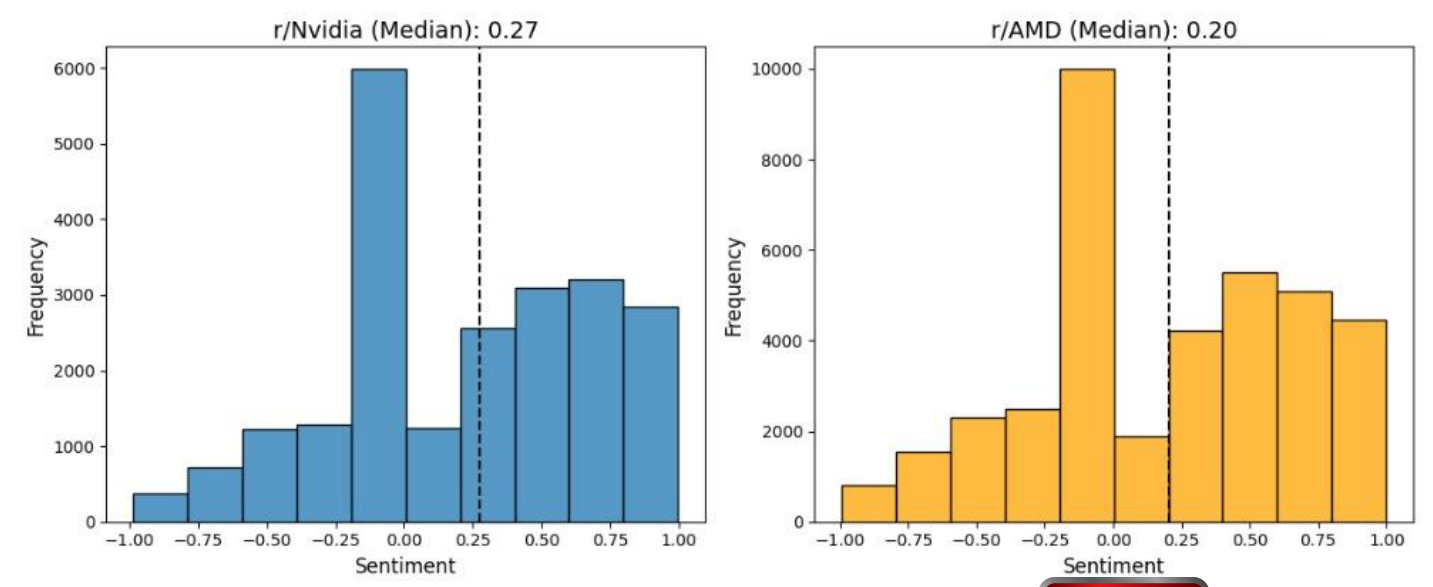
Sentiment Analysis

05

Sentiment Analysis

- What is sentiment analysis?
  - Process of determining if a piece of writing is positive, negative or neutral.
  - Natural language processing(NLP) technique to understand people's emotion and opinions towards certain products.
- Method used:
  - VADER's SentimentIntensityAnalyzer()

Sentiment Showdown



Sentiment Showdown

- Scores:
  - AMD has a positive sentiment score
  - nVidia still ranks higher in sentiment score

- Percentage of negative posts
  - 

    % of negative posts 15.2%
  - 

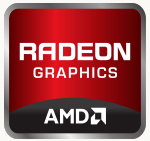
    % of negative posts 10.9%

- For every n comments on r/amd, 14.4% more negative comments than on r/nVidia

Conclusion

Problem Statement

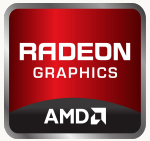
It is unclear whether the **perception of AMD** products being more problematic is a **marketing** or an **engineering issue**. Some light can be shed on this through sentiment analysis.



Conclusion

Model	Train Score	Test Score	Precision	Recall	F1-Score	Fit time
Naive Bayes with cvec	0.990	0.944	0.909	0.820	0.862	18s
Naive Bayes with tvec	0.993	0.941	0.915	0.739	0.818	14s
Random Forest with tvec	0.999	0.937	0.932	0.736	0.823	43s
Adaboost with tvec	0.999	0.913	0.885	0.735	0.819	58s
Sequential Neural Net with tvec	N/A	0.865	0.931	0.752	0.832	5s (CUDA)

- Simple sequential neural network is overall the best model.
- however, it is dependent on nVidia proprietary technology.



Conclusion

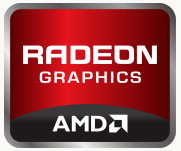
- Best model based on **precision and time for duration fit**:

- **Random Forest with TFIDF Vectorizer**
  - slightly lower precision score
  - Short time to fit model

- Neural Network has the best precision score, but can only be run on Nvidia GPU

Best model for deployment in other forums
Random Forest with TFIDF Vectorizer

Recommendations

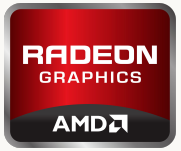


Model

The concern is they are still few non-AMD posts as this model is used to pick out AMD-related posts from GPU forums.

To re-trained model as a multi-classification model to take into Intel Arc GPUs into account once they achieve further market penetration

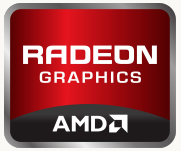
Recommendations



Sentiment Analysis

- Engineering - development for greater driver stability
- Development and implementation of comprehensive testing suite before new releases
- Differences between AMD and nVidia
  - Lack of settings (AMD GPUs)
  - Having more available parameters (so users have the option to tune their GPUs for their system)

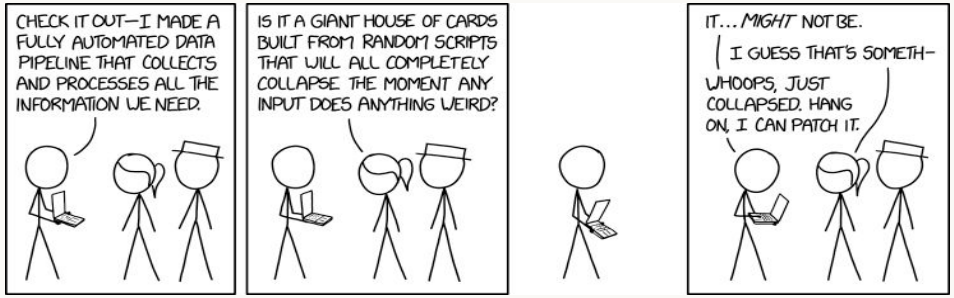
Constraints



Price Related Complaints

- Price related complaints of Lovelace generation GPUs
- AMD CPU posts (which have been extremely well received) have not been filtered out.
- likely an even greater difference in technical complaints ~38.9%.

THANK YOU



RESOURCES

Reddit

- Data scrapped from subreddits AMD and Nvidia