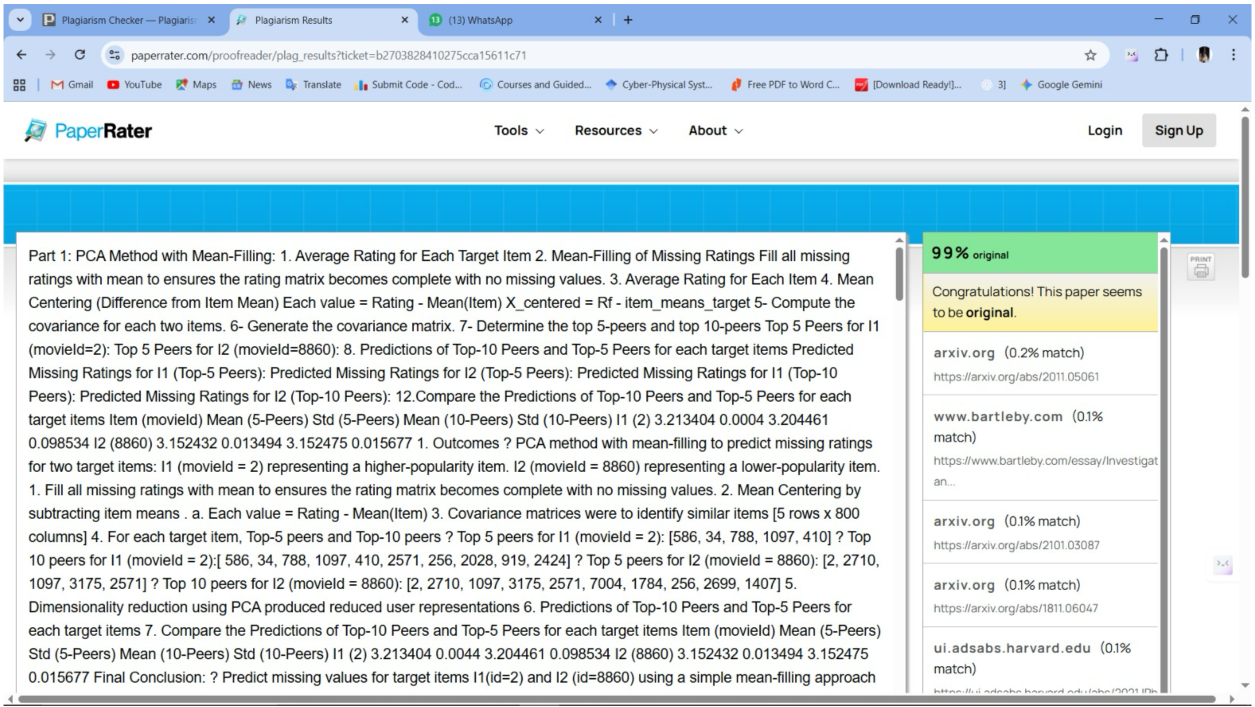
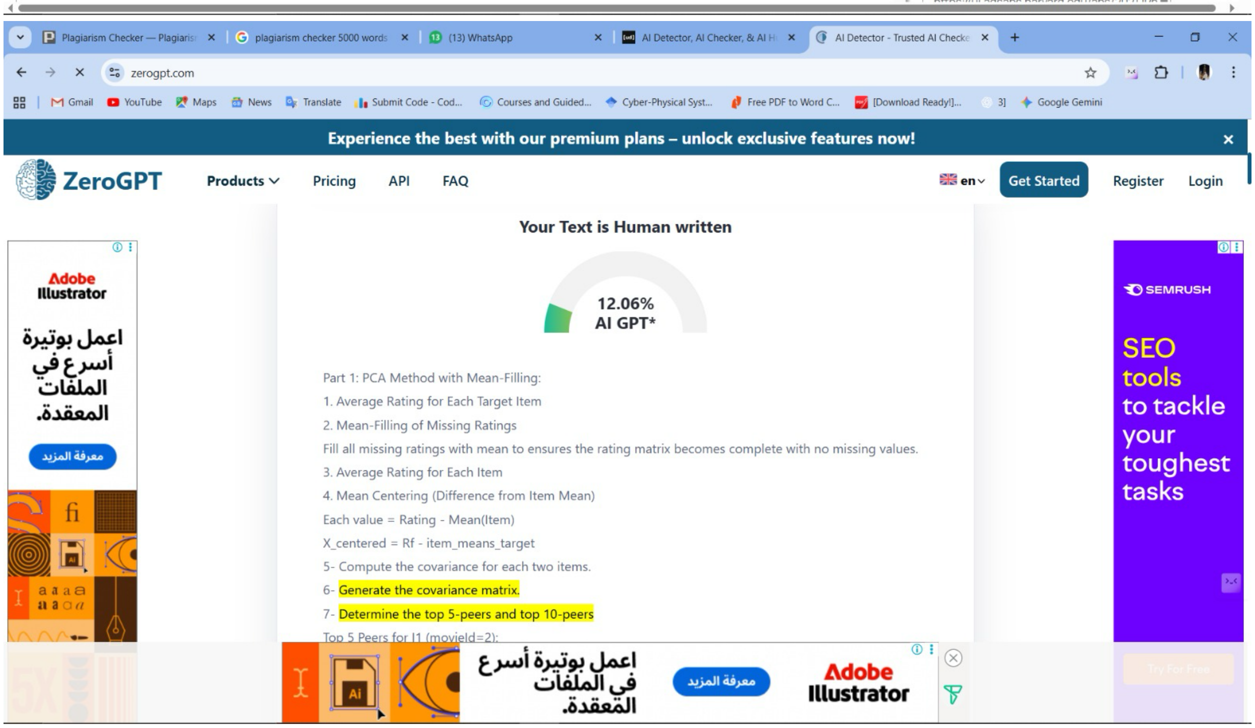
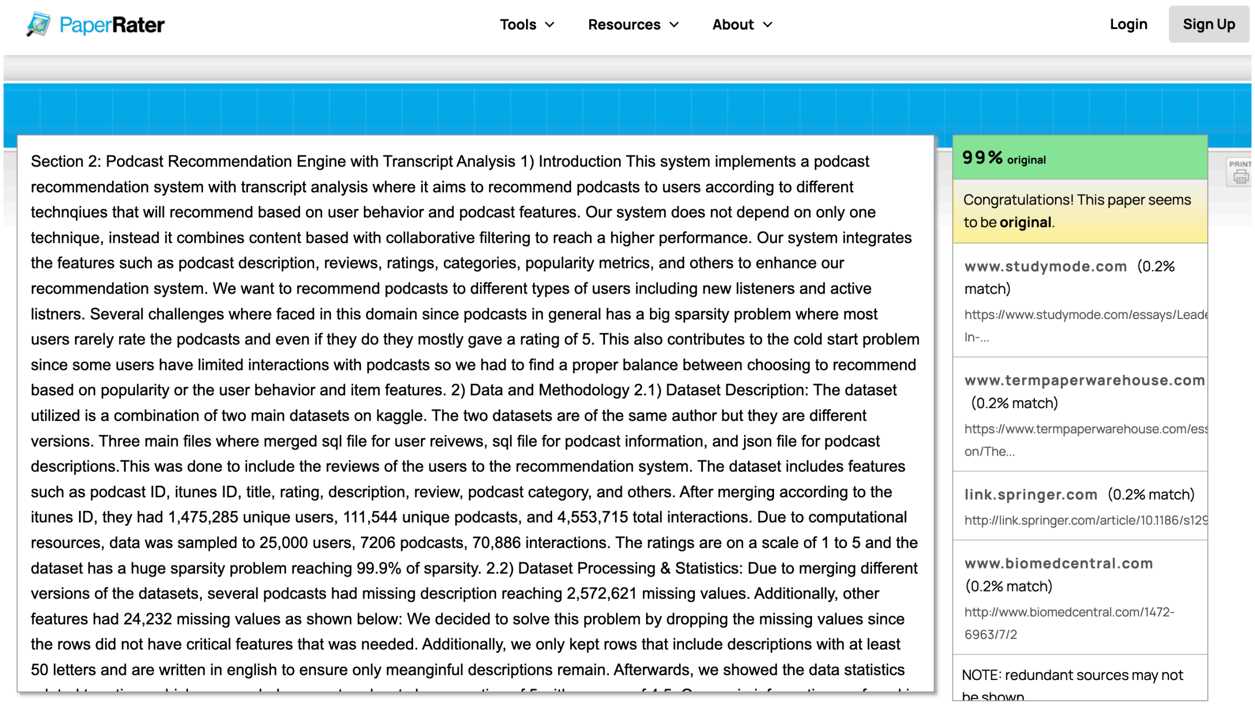
Plagiarism Report

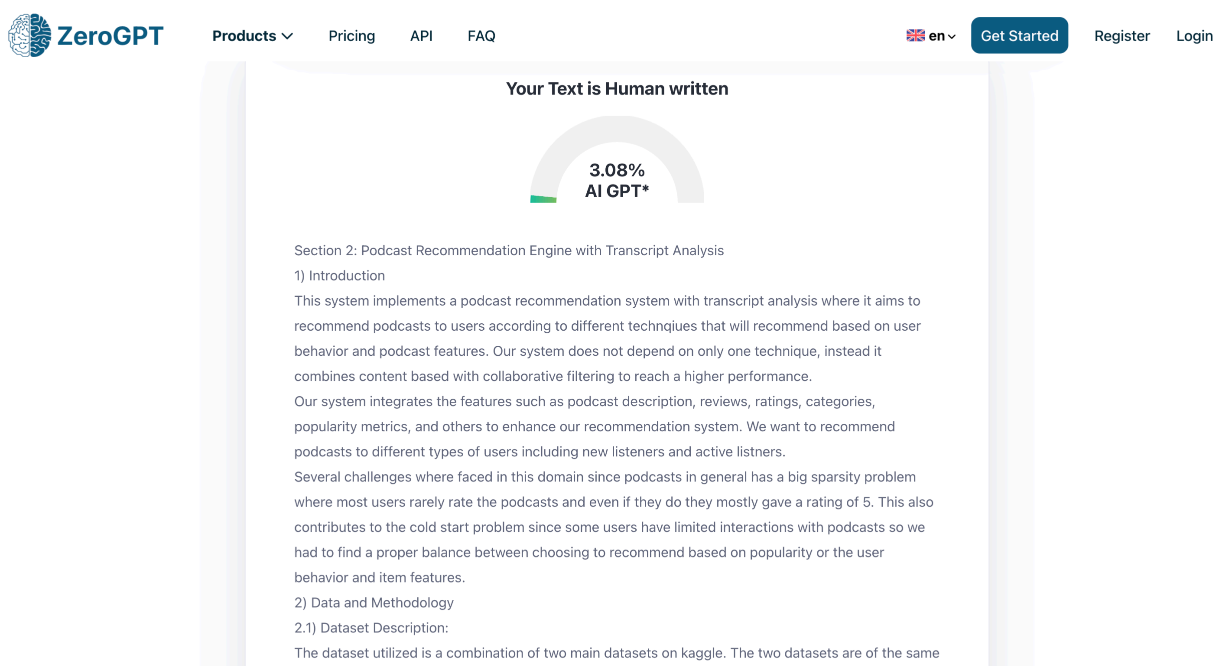
1) Section 1:





2) Section 2:


PaperRater
Tools
Resources
About
Login
Sign Up
Section 2: Podcast Recommendation Engine with Transcript Analysis 1) Introduction This system implements a podcast recommendation system with transcript analysis where it aims to recommend podcasts to users according to different technqiues that will recommend based on user behavior and podcast features. Our system does not depend on only one technique, instead it combines content based with collaborative filtering to reach a higher performance. Our system integrates the features such as podcast description, reviews, ratings, categories, popularity metrics, and others to enhance our recommendation system. We want to recommend podcasts to different types of users including new listeners and active listners. Several challenges where faced in this domain since podcasts in general has a big sparsity problem where most users rarely rate the podcasts and even if they do they mostly gave a rating of 5. This also contributes to the cold start problem since some users have limited interactions with podcasts so we had to find a proper balance between choosing to recommend based on popularity or the user behavior and item features. 2) Data and Methodology 2.1) Dataset Description: The dataset utilized is a combination of two main datasets on kaggle. The two datasets are of the same author but they are different versions. Three main files where merged sql file for user reivews, sql file for podcast information, and json file for podcast descriptions.This was done to include the reviews of the users to the recommendation system. The dataset includes features such as podcast ID, itunes ID, title, rating, description, review, podcast category, and others. After merging according to the itunes ID, they had 1,475,285 unique users, 111,544 unique podcasts, and 4,553,715 total interactions. Due to computational resources, data was sampled to 25,000 users, 7206 podcasts, 70,886 interactions. The ratings are on a scale of 1 to 5 and the dataset has a huge sparsity problem reaching 99.9% of sparsity. 2.2) Dataset Processing & Statistics: Due to merging different versions of the datasets, several podcasts had missing description reaching 2,572,621 missing values. Additionally, other features had 24,232 missing values as shown below: We decided to solve this problem by dropping the missing values since the rows did not have critical features that was needed. Additionally, we only kept rows that include descriptions with at least 50 letters and are written in english to ensure only meanginful descriptions remain. Afterwards, we showed the data statistics
99% original
Congratulations! This paper seems to be original.
www.studymode.com (0.2% match)
https://www.studymode.com/essays/Leade
In-...
www.termpaperwarehouse.com (0.2% match)
https://www.termpaperwarehouse.com/ess
on/The...
link.springer.com (0.2% match)
http://link.springer.com/article/10.1186/s129
www.biomedcentral.com (0.2% match)
http://www.biomedcentral.com/1472-6963/7/2
NOTE: redundant sources may not


ZeroGPT
Products
Pricing
API
FAQ
en
Get Started
Register
Login
Your Text is Human written
3.08%
AI GPT*
Section 2: Podcast Recommendation Engine with Transcript Analysis
1) Introduction
This system implements a podcast recommendation system with transcript analysis where it aims to recommend podcasts to users according to different technqiues that will recommend based on user behavior and podcast features. Our system does not depend on only one technique, instead it combines content based with collaborative filtering to reach a higher performance.
Our system integrates the features such as podcast description, reviews, ratings, categories, popularity metrics, and others to enhance our recommendation system. We want to recommend podcasts to different types of users including new listeners and active listners.
Several challenges where faced in this domain since podcasts in general has a big sparsity problem where most users rarely rate the podcasts and even if they do they mostly gave a rating of 5. This also contributes to the cold start problem since some users have limited interactions with podcasts so we had to find a proper balance between choosing to recommend based on popularity or the user behavior and item features.
2) Data and Methodology
2.1) Dataset Description:
The dataset utilized is a combination of two main datasets on kaggle. The two datasets are of the same