**Introduction**:

This study investigates the potential correlation between IMDb user ratings and the popularity of films on Ekşi Sözlük, a widely used Turkish social platform. Two main data sources were used for this analysis. The first one was obtained from (https://datasets.imdbws.com), IMDb's official datasets. Particularly, `title.basics.tsv.gz` was used to extract fundamental information such as film titles, release years, and types, while `title.ratings.tsv.gz` provided user ratings. The second was collected (https://eksisozluk.com), which includes entries that discussions related to films from the users. The number of entries for each film on Ekşi Sözlük was collected to represent the level of public interest. By combining these two datasets, this study investigates whether IMDb ratings are associated with a film’s popularity, as measured by entry counts on Ekşi Sözlük.

**EDA Results and Interpretations**

The exploratory data analysis was conducted across four distinct time periods: 2008–2011, 2012–2015, 2016–2019, and 2020–2024. In each period, the relationship between IMDb scores and Ekşi Sözlük entry counts was visualized and analyzed using scatter plots and linear regression trend lines. Pearson correlation coefficients and p-values were also computed to quantify the strength and statistical significance of the relationships.

**🔹 2008–2011**

* **Pearson Correlation**: 0.346
* **P-value**: 0.0489
* metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

  Yapay zeka tarafından oluşturulan içerik yanlış olabilir.This period shows a moderate positive correlation between the number of entries and IMDb scores. The p-value is just below the 0.05 threshold, indicating statistical significance. This suggests that films with higher IMDb scores tended to attract more release attention on Ekşi Sözlük during this period.

**🔹2012–2015**

* **Pearson Correlation**: 0.313
* **P-value**: 0.0626

metin, çizgi, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

* A moderate positive correlation is again observed, but the p-value is slightly above 0.05, indicating marginal non-significance. This suggests a possible trend, though not statistically confirmed, between higher IMDb scores and increased entry counts.

**🔹 2016–2019**

* **Pearson Correlation**: 0.335
* **P-value**: 0.0878

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, metin, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

* The correlation remains positive but relatively weak and statistically insignificant. The dispersion of points indicates that other factors may have played a larger role in entry frequency during this period.

**🔹 2020–2024**

* **Pearson Correlation**: 0.152
* **P-value**: 0.4135

metin, çizgi, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

* This period exhibits the weakest correlation among all four. The high p-value indicates that there is no statistically significant relationship between IMDb scores and popularity on Ekşi Sözlük during recent years. Increased social media fragmentation or changing audience behavior might explain this pattern.

**Hypothesis Testing**

To assess whether there is a statistically significant link between IMDb scores entry counts on Ekşi Sözlük, Pearson correlation analysis was performed for each time period. The results indicate a moderate positive correlation for the years 2008–2011 (r = 0.346, p = 0.0489), which is statistically significant at the 0.05 level. However, in the later periods, although weak to moderate correlations were observed, the p-values exceeded 0.05, suggesting no significant linear relationship in those cases. Finally, the hypothesis that higher IMDb scores are associated with greater entry counts is only supported for the 2008–2011 interval.

**Machine Learning Methods**

We evaluated two machine learning models **Linear Regression** and **KNN Regressor** (k=3) on datasets from the years 2008-2011, 2012-2015, 2016-2019, and 2020-2024 to predict IMDb scores based on the estimated Ekşi Sözlük entry counts.

In 2008-2011, Linear Regression slightly outperformed KNN, though both models performed poorly, suggesting a weak relationship between entry counts and IMDb scores. In 2012-2015, KNN showed better predictive power with an R² of 0.195, outperforming Linear Regression’s 0.094.

The best overall results were observed in 2016-2019 period, where both models had the lowest MSE and highest R² scores. KNN again outperformed Linear Regression with an R² of 0.242, indicating a moderately strong nonlinear relationship.

However, in 2020-2024, both models performed poorly, with KNN yielding a very low and even negative R² (−0.841), suggesting that the entry count was not a reliable predictor of IMDb score that year.

Overall, **KNN Regressor performed better in years where nonlinear relationships were more prominent**, while **Linear Regression remained more consistent** across all years. These findings suggest that model performance is highly dependent on the structure and quality of the data for each year.

**Challenges Encountered**

During the data collection process, several challenges related to film title matching were encountered. For example, the film *Roma* by Alfonso Cuarón is listed on Ekşi Sözlük under the title "Roma (Alfonso Cuarón filmi)", which differs from its standard IMDb title. This inconsistency required manual verification for accurate mapping. Additionally, certain films such as *Her(2013)* were excluded from the analysis due to ambiguity in the search results. The keyword "her" produced a wide range of unrelated entries, making it difficult to isolate data specific to the film.

**General Insight**

Across all four periods, a weak-to-moderate positive correlation is observed between IMDb ratings and entry counts. However, the strength and significance of this relationship declines over time, suggesting that IMDb scores have become less predictive of Ekşi Sözlük engagement in recent years.