To understand why **Alien** is ranked higher than **Aliens** in your recommendation system, we need to analyze the **combined features** used for calculating similarity. The similarity score is based on the **TF-IDF vectorization** of the combined features (genres, keywords, tagline, cast, and director). Let’s break it down step by step:

**1. Combined Features for Each Movie**

Here are the combined features for **Avatar**, **Alien**, and **Aliens** based on your dataset:

**Avatar**

Copy

Action Adventure Fantasy Science Fiction culture clash future space war space colony society In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. Sam Worthington Zoe Saldana Sigourney Weaver Stephen Lang Michelle Rodriguez James Cameron

**Alien**

Copy

Horror Action Thriller Science Fiction android countdown space marine space suit beheading During its return to the earth, commercial spaceship Nostromo intercepts a distress signal from a distant planet. When a three-member team of the crew discovers a chamber containing thousands of eggs on the planet, a creature inside one of the eggs attacks an explorer. The entire crew is unaware of the impending nightmare set to descend upon them when the alien parasite planted inside its unfortunate host is birthed. Tom Skerritt Sigourney Weaver Veronica Cartwright Harry Dean Stanton John Hurt Ridley Scott

**Aliens**

Copy

Horror Action Thriller Science Fiction android extraterrestrial technology space marine spaceman cryogenics When Ripley's lifepod is found by a salvage crew over 50 years later, she finds that terra-formers are on the very planet they found the alien species. When the company sends a family of colonists out to investigate her story, all contact is lost with the planet and colonists. They enlist Ripley and the colonial marines to return and search for answers. Sigourney Weaver Michael Biehn James Remar Paul Reiser Lance Henriksen James Cameron

**2. Key Observations**

1. **Director**:
   * **Avatar** and **Aliens** share the same director (**James Cameron**).
   * **Alien** is directed by **Ridley Scott**.
2. **Genres**:
   * **Avatar**: Action, Adventure, Fantasy, Science Fiction
   * **Alien**: Horror, Action, Thriller, Science Fiction
   * **Aliens**: Horror, Action, Thriller, Science Fiction

Both **Alien** and **Aliens** share the same genres, but **Avatar** has Adventure and Fantasy instead of Horror and Thriller.

1. **Keywords**:
   * **Avatar**: culture clash, future, space war, space colony, society
   * **Alien**: android, countdown, space marine, space suit, beheading
   * **Aliens**: android, extraterrestrial, technology, space marine, spaceman, cryogenics

**Alien** and **Aliens** share keywords like android and space marine, but **Alien** has countdown and beheading, which might align more closely with **Avatar**'s themes of conflict and survival.

1. **Tagline**:
   * **Avatar**: Enter the World of Pandora.
   * **Alien**: In space no one can hear you scream.
   * **Aliens**: This Time It's War

The tagline of **Alien** (In space no one can hear you scream) might resonate more with **Avatar**'s themes of isolation and danger in space.

1. **Cast**:
   * **Avatar** and **Aliens** both feature **Sigourney Weaver**, which could increase their similarity.
   * However, **Alien** also features **Sigourney Weaver**, so this doesn’t explain the difference in ranking.

**3. Why Alien is Ranked Higher**

The **TF-IDF vectorizer** assigns weights to words based on their importance in the dataset. Here’s why **Alien** might have a higher similarity score than **Aliens**:

1. **Keyword Overlap**:
   * **Alien** has keywords like countdown and beheading, which might align more closely with **Avatar**'s themes of conflict and survival.
   * **Aliens** has keywords like extraterrestrial and cryogenics, which are less directly related to **Avatar**'s themes.
2. **Tagline**:
   * The tagline of **Alien** (In space no one can hear you scream) might be more similar to **Avatar**'s themes of isolation and danger in space compared to **Aliens**' tagline (This Time It's War).
3. **TF-IDF Weighting**:
   * Words like space marine and android are common in both **Alien** and **Aliens**, but **Alien** might have additional unique words (e.g., countdown, beheading) that align better with **Avatar**'s features.
4. **Genre Overlap**:
   * While both **Alien** and **Aliens** share genres with **Avatar**, **Alien**'s Horror and Thriller genres might align more closely with **Avatar**'s themes of danger and survival.

**4. Debugging the Similarity**

To confirm this, you can calculate the **cosine similarity** between **Avatar** and **Alien** vs. **Avatar** and **Aliens**:

python

Copy

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.metrics.pairwise import cosine\_similarity

# Combined features for Avatar, Alien, and Aliens

avatar\_features = "Action Adventure Fantasy Science Fiction culture clash future space war space colony society In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. Sam Worthington Zoe Saldana Sigourney Weaver Stephen Lang Michelle Rodriguez James Cameron"

alien\_features = "Horror Action Thriller Science Fiction android countdown space marine space suit beheading During its return to the earth, commercial spaceship Nostromo intercepts a distress signal from a distant planet. When a three-member team of the crew discovers a chamber containing thousands of eggs on the planet, a creature inside one of the eggs attacks an explorer. The entire crew is unaware of the impending nightmare set to descend upon them when the alien parasite planted inside its unfortunate host is birthed. Tom Skerritt Sigourney Weaver Veronica Cartwright Harry Dean Stanton John Hurt Ridley Scott"

aliens\_features = "Horror Action Thriller Science Fiction android extraterrestrial technology space marine spaceman cryogenics When Ripley's lifepod is found by a salvage crew over 50 years later, she finds that terra-formers are on the very planet they found the alien species. When the company sends a family of colonists out to investigate her story, all contact is lost with the planet and colonists. They enlist Ripley and the colonial marines to return and search for answers. Sigourney Weaver Michael Biehn James Remar Paul Reiser Lance Henriksen James Cameron"

# Create TF-IDF vectors

vectorizer = TfidfVectorizer()

tfidf\_matrix = vectorizer.fit\_transform([avatar\_features, alien\_features, aliens\_features])

# Calculate cosine similarity

similarity\_avatar\_alien = cosine\_similarity(tfidf\_matrix[0:1], tfidf\_matrix[1:2])[0][0]

similarity\_avatar\_aliens = cosine\_similarity(tfidf\_matrix[0:1], tfidf\_matrix[2:3])[0][0]

print("Similarity (Avatar, Alien):", similarity\_avatar\_alien)

print("Similarity (Avatar, Aliens):", similarity\_avatar\_aliens)

**5. Possible Fixes**

If you want **Aliens** to rank higher than **Alien**, you can:

* **Increase the Weight of Director**: Since both **Avatar** and **Aliens** are directed by James Cameron, you can give more weight to the director feature in the combined features.
* **Adjust Keywords**: Ensure the keywords for **Aliens** are more aligned with **Avatar**'s themes.

**Conclusion**

The higher similarity score for **Alien** is likely due to its **keywords** and **tagline** aligning more closely with **Avatar**'s themes. By inspecting the combined features and TF-IDF vectors, you can better understand the ranking and make adjustments if needed.

Let me know if you need further clarification or help! 😊