Did you describe the dataset, and any challenging characteristics it has?

Did you describe (briefly) the data cleaning and preprocessing?

Did you describe the considerations that guided your design of learning systems?

Did you describe (briefly) various methods you tried and the results you obtained?

Did you describe the learning system you ended up using?

Did you provide a prediction (and explanation) of the generalization error you expect your system to have?

Explain your solution, describe your work process, and do your best to justify any design decisions you have made. Feel free to include supporting figures if you want

**data cleaning and preprocessing**

החלטנו למחוק את הפיצ'ר של spoken\_languages כי לדעתנו הוא לא נותן ערך מוסף משמעותי לoriginal\_language. את ה original\_languageחילקנו לאינדיקטורים כי אין ביניהם יחס סדר.

פיצ'ר נוסף שחילקנו לאינדיקטורים (dummy vectors) הוא הז'אנר של הסרט.

אין משמעות לת"ז שנתנו ולכן הורדנו את הפיצ'ר הזה, כמו כן, דף הבית של הסרט ושם הסרט אינם מוסיפים מידע חיוני והם ייחודים לכל סרט.

belongs\_to\_collection – האם סרט הוא חלק מסדרה של סרטים יכול להוסיף לנו מידע, אך בצורתו הגולמית הוא לא הוסיף ולכן אם סרט אינו חלק מסדרה נתנו את הערך 0 ואם הוא כן נתנו את הערך 1.

**the learning system**

**the generalization error**