**שיבוץ שחקנים ראשיים לסרט תוך   
מקסום רווחים מהסרט**

**פרויקט בבינה מלאכותית**

**מגישות:**

חן חבקוק ת"ז 312320997

טל עובד ת"ז 206129710

**מנחה:**

גיא קושילביץ

**הקדמה**

**רקע**

תעשיית הסרטים

קולנוע הוא תחום [אמנות](https://he.wikipedia.org/wiki/%D7%90%D7%9E%D7%A0%D7%95%D7%AA) העוסק ביצירתם ובהקרנתם של [סרטי קולנוע](https://he.wikipedia.org/wiki/%D7%A1%D7%A8%D7%98_%D7%A7%D7%95%D7%9C%D7%A0%D7%95%D7%A2).   
הקולנוע הוא מערכת תרבותית, במסגרתה נוצרים מאות ואלפי תוצרי תרבות הנצפים על ידי מיליונים ברחבי העולם. הקולנוע הוא בין השאר כלי [תקשורת המונים](https://he.wikipedia.org/wiki/%D7%AA%D7%A7%D7%A9%D7%95%D7%A8%D7%AA_%D7%94%D7%9E%D7%95%D7%A0%D7%99%D7%9D), ובעל השפעות גדולות על החברה והתרבות.   
תעשיית הקולנוע והסרטים היא המצרך האמנותי הפופולרי ביותר.

המילה "קולנוע" היא הלחם של המילים [קול](https://he.wikipedia.org/wiki/%D7%A7%D7%95%D7%9C) ו[תנועה](https://he.wikipedia.org/wiki/%D7%AA%D7%A0%D7%95%D7%A2%D7%94_(%D7%A4%D7%99%D7%96%D7%99%D7%A7%D7%94)) כאשר הדגש הוא על המילה הקול .   
מקורו בהתפתחות הטכנולוגית - הקולנוע בא לאחר ה[ראינוע](https://he.wikipedia.org/wiki/%D7%A8%D7%90%D7%99%D7%A0%D7%95%D7%A2) - הסרט האילם.

בזמן ה[ראינוע](https://he.wikipedia.org/wiki/%D7%A8%D7%90%D7%99%D7%A0%D7%95%D7%A2), הבדלי שפה שבין ארץ לאחרת לא היוו מגבלה מכיוון שהסרטים היו אילמים. מדי פעם, כשנוצר צורך, היו מקרינים במקביל לסרט כותרות (במעין שלט נלווה) וכשרצו להקרין סרט בארץ אחרת, החליפו את הכותרות לשפה המדוברת באותה ארץ.

הפקת סרטים היא יקרה מאד לכן ההשקעות הגדולות מתבצעות לרוב על ידי אנשים בעלי ניסיון. הסרטים מדורגים לרמות שונות, כשרמות גבוהות יותר פירושן השקעה גדולה יותר ויקרה יותר.

הקולנוע ההוליוודי

הקולנוע ה[הוליוודי](https://he.wikipedia.org/wiki/%D7%94%D7%95%D7%9C%D7%99%D7%95%D7%95%D7%93) החל את דרכו יחד עם הקולנוע עצמו.   
הוליווד עצמה הוקמה בתחילת שנות העשרים, כשיוצרי הראינוע נדדו מערבה. שם רכשה קבוצה מהם אדמה מבעליה, ששמם היה הוליווד, והחליטו להשאיר את השם. הוליווד הייתה במקום מושלם למגוון של לוקיישנים שונים. החל מהנוף ההררי מצפון, הים במערב, ההרים והמדבר במזרח, ועמקי קליפורניה הפוריים.

הקולנוע זכה לדחיפה גדולה בשל יצירתו של הקולנוען האמריקני ד.וו.גריפית שהתחיל את הקריירה הקולנועית שלו בתור שחקן, אך מהר מאוד התחיל גם לצלם ותוך זמן קצר גם קיבל הזדמנות לביים.   
בשנת[1915](https://he.wikipedia.org/wiki/1915)  ביים את סרטו הראשון "לידתה של אומה", סרט שיצר למעשה את שפת הקולנוע כפי שאנו מכירים אותה היום כמו: עריכה בתוך הסצנה, שימוש ב[קלוז-אפ](https://he.wikipedia.org/wiki/%D7%A7%D7%9C%D7%95%D7%96_%D7%90%D7%A4) וקצב עריכה שונה בהתאם לדרמטיות שבסצנה.   
סרטו השני, "אי סובלנות" כלל סצנות המונים ותפאורות בגובה עשרות מטרים, מצלמות שהועלו עם מעליות והמצאות רבות שהפכו את הסרט, שאורכו 163 דקות, לתצוגת איכות קולנועית כבר ב-1916.

קולנוען נוסף וידוע שהביא את הראינוע לפסגות חדשות היה [צ'ארלי צ'פלין](https://he.wikipedia.org/wiki/%D7%A6%27%D7%90%D7%A8%D7%9C%D7%99_%D7%A6%27%D7%A4%D7%9C%D7%99%D7%9F), אשר נחשב לאחת מהדמויות החשובות והמשפיעות ביותר בהיסטוריה של [תעשיית הקולנוע](https://he.wikipedia.org/wiki/%D7%AA%D7%A2%D7%A9%D7%99%D7%99%D7%AA_%D7%94%D7%A7%D7%95%D7%9C%D7%A0%D7%95%D7%A2). סרטיו של צ'פלין הביאו לעולם קולנוע חדש שלא נראה כמוהו, כוריאוגרפיה ייחודית ושימוש מבריק באביזרים ותפאורה.

תור הזהב של הוליווד

הוליווד ידועה כמרכז הפקות הסרטים העולמי. האולפן הראשון בהוליווד הוקם ב1911.  
בשנות השלושים ובשנות הארבעים, במהלך התקופה אשר זכתה לכינוי "עידן הזהב של הוליווד", הפיקו אולפני הסרטים בהוליווד בעיקר סרטי [מערבונים](https://he.wikipedia.org/wiki/%D7%9E%D7%A2%D7%A8%D7%91%D7%95%D7%9F), סרטי סלפסטיק (קומדיה משולבת בתנועות גופניות מכאיבות ומרעישות), סרטי "פילם נואר" (סרטי פשע, בעלי אווירה אפלה), סרטים מוזיקליים (מחזמר), סרטי אנימציה מצוירים ו[סרטים ביוגרפיים](https://he.wikipedia.org/wiki/%D7%A1%D7%A8%D7%98_%D7%91%D7%99%D7%95%D7%92%D7%A8%D7%A4%D7%99). באותה העת אלפי אנשים הועסקו על ידי אולפני הסרטים - שחקנים, מפיקים, במאים, כותבים, פעלולנים, אנשי מלאכה וטכנאים. בבעלותם של אולפני הסרטים היו מאות בתי קולנוע בערים שונות ברחבי ארצות הברית בהם הוצגו הסרטים שלהם. נקודת השיא של תעשיית הסרטים האמריקאית הייתה במהלך אמצע [שנות הארבעים](https://he.wikipedia.org/wiki/%D7%A9%D7%A0%D7%95%D7%AA_%D7%94-40_%D7%A9%D7%9C_%D7%94%D7%9E%D7%90%D7%94_%D7%94-20) כאשר אולפני הקולנוע הפיקו באותה העת כ-400 סרטים בשנה אשר הוצגו בפני קהל של 90 מיליון צופים אמריקאים מדי שבוע.

בשנת 1941 מוציא לאקרנים הבמאי [אורסון ולס](https://he.wikipedia.org/wiki/%D7%90%D7%95%D7%A8%D7%A1%D7%95%D7%9F_%D7%95%D7%9C%D7%A1) את הסרט [האזרח קיין](https://he.wikipedia.org/wiki/%D7%94%D7%90%D7%96%D7%A8%D7%97_%D7%A7%D7%99%D7%99%D7%9F) - סרט מופתי שנחשב לאחד החשובים בהיסטוריה של הקולנוע. חידושיו כללו שפה קולנועית עשירה, שבירה של מסגרת הזמנים המוכרת ובנייה של דמויות מורכבות ובעלות מספר צדדים.

לאורך השנים, התעשייה התפתחה, צברה ניסיון ושידרגה את היכולות הטכנולוגיות, המקצועיות והאומנתיות.

**רקע מהקורס**

ראינו בקורס כי תהליך החיזוי הוא תוצאה של תהליכי למידה שונים.  
למידה הינה תהליך, המקבל התנסויות כקלט, ומבצע שינויים בבסיס ידע במטרה לשפר, על פי מדד נתון, את היכולת של פותר הבעיות, המשתמש בבסיס הידע לפתור קבוצת בעיות.  
שלבי הלמידה המרכזיים:  
1. מידול הבעיה  
2. איסוף דוגמאות וייצוגן  
3. עיבוד מידע  
4. שימוש באלגוריתם למידה – מסווג  
5. הערכת ביצועים

קיימות בעיות שונות, כמו בעיות בינאריות שמטרתן לקבוע האם אובייקט שייך לקבוצה או לא, ובעיות רגרסיה שמטרתן לנבא מספר ממשי (כפי שנתעסק בעבודתנו).

**הרעיון הכללי של הפרויקט (תיאור הבעיה):**

בעידן של היום, לאחר התפתחות רבה של תעשיית הסרטים, חשוב לתכנן כל סרט מראש במטרה למקסם את הרווחים שניתן להפיק ממנו.

ולכן, המטרה העיקרית בפרויקט: בהינתן סרט יחד עם תכונות (ז'אנר, במאי, תסריטאי, תקציב סרט..) המאפיינות אותו וכמות שחקנים נדרשת, נרצה לייצר מערכת המחזירה את שילוב השחקנים האופטימאלי כך שרווחי הסרט בקופות בתי הקולנוע ((box office יהיו מקסימאליים.

**דרך הפתרון**

את הבעיה שהצגנו החלטנו לפתור בשני שלבים מרכזיים:

1. מסווג -   
   המסווג יקבל כקלט סרט יחד עם התכונות הרלוונטיות המאפיינות את הסרט ויחזיר את הצפי לרווח הסרט בקופות בתי הקולנוע.
2. אלגוריתם -   
   האלגוריתם יקבל סרט עם תכונות המאפיינות אותו, וכמות שחקנים נדרשת ויחזיר את קבוצת השחקנים עבורה צפי ההכנסות מהסרט הוא מקסימאלי.

כעת נפרט:

שלב א' – בניית המסווג/רגרסור

1. השגת הנתונים

מאגר הנתונים הוא המידע עליו מתבסס כל הפרוייקט ולכן חשוב מאוד שיכיל נתונים טובים, ובנוסף שיכיל כמה שיותר סרטים – מה שיגדיל את רמת הדיוק המתבקשת.

ולכן, בשלב הראשון חיפשנו מאגרי נתונים מספיק גדולים שיכילו כמה שיותר תכונות שנראו לנו רלוונטיות לקביעת הרווחים העתידיים של סרט.   
לאחר חיפוש נרחב ברשת בחרנו בשני מאגרי נתונים מאתר Kaggle, שהכילו תכונות שונות ואיחדנו אותם למאגר נתונים אחד.   
בשלב זה מאגר הנתונים הכיל כ4800 סרטים שונים עם תכונות כגון:

* שפת מקור
* פופולריות הסרט
* חברת הפקות
* תאריך יציאה של הסרט
* רווחי הסרט
* אורך הסרט (זמן)
* ממוצע הציון שקיבל הסרט
* כותרת הסרט
* שחקנים
* הצוות (במאי, מפיק, תסריטאים...)
* תקציב הסרט
* ז'אנרים
* אתר הסרט
* מילות מפתח

בשלב זה עברנו על מאגר הנתונים, וידאנו בצורה רנדומלית חלק מהסרטים לאמת את המידע וניסינו להבין כמה מידע יהיה עלינו להשלים.  
בעיות שנתקלנו בהן בחלק זה:

* 1. מציאת מאגר נתונים עם כמות גדולה של סרטים שיכיל כמה שיותר נתונים שלדעתנו יעזרו לחזות את רווחי הסרט – קשה למצוא מאגר יחיד שיכיל את "כל מה שאנחנו צריכים" ולכן בסוף מצאנו שני מאגרים שהיה ניתן לשלב אותם יחסית בקלות וביצענו את השילוב.
  2. מציאת מאגר נתונים עם כמה שפחות "חוסרי מידע" – במעבר ידני מאוד קשה לראות בכמה חוסרים מדובר, נרחיב על ההתמודדות עם שלב זה בחלק הבא.

1. קרצוף נתונים

מאגר הנתונים שהיה בידינו בשלב זה היה רחוק מלהיות מושלם לעבודה עם מסווג/רגרסור.   
לכן בשלב זה התחלנו לחפש ולכתוב קטעי קוד שיעזרו לנו להשלים את המידע החסר או לתקן מידע קיים (למשל – שמות של סרטים/שחקנים שהכילו תווים בלתי קריאים).

דברים שביצענו:

* השלמת ערכים חסרים של קרוב ל1000 סרטים של תקציב ורווח.
* נרמול ערכי התקציב, הרווח.
* בניית מאגרי שחקנים, במאים, מפיקים ותסריטאים מהמידע הקיים.
* בכל מאגר, נתינת "ציון" מספרי לכל בעל מקצוע. ציון המהווה ממוצע כל הרווחים של הסרטים בהם אותו בעל מקצוע השתתף.
* לבסוף איחדנו את כל המאגרים לכדי DB המכיל עבור כל סרט את:  
  שמו, "ציון" (המוסבר לעיל) עבור השחקנים, הבמאים, המפיקים והכותבים, תקציב מנורמל ורווח מנורמל.

1. בחירת אלגוריתם
   1. למה מתאים לבעיה + הסבר האלגוריתם
2. אימון המודל
   1. תוצאות, טבלאות
3. בחינת המודל

שלב ב' - אלגוריתם

1. בדיקה מיטבית של שלשות שחקנים

סיכום

**ביבליוגרפיה**

* המושג [קולנוע](https://he.wikipedia.org/wiki/%D7%A7%D7%95%D7%9C%D7%A0%D7%95%D7%A2) (ויקיפדיה).
* מאגר נתונים הנתון כקובץ .csv של 4800 סרטים ומאפיינים עליהם כולל שחקנים הלקוח מאתר Kaggle - <https://www.kaggle.com/tmdb/tmdb-movie-metadata/data>
* מאגר נתונים הנתון כקובץ .csvשל 2500 סרטים ומאפיינים עליהם כולל ז'אנר, חברה מפיצה, רווחים (Box office) הלקוח מאתר Kaggle - <https://www.kaggle.com/yjeong5126/box-office-data-20172019>
* השלמת מידע חסר למאגר הנתונים – ויקיפדיה, IMDB.