**דוח מסקנות: בחירת אלגוריתם העלילה המתאים ביותר ליצירת שלדי סיפורים**

**מוטיבציה לניסוי**

כחלק ממטרות הפרויקט, התבצע ניסוי לבחירת אלגוריתם עבור פיתוח שלד העלילה הטוב ביותר שיספק בסיס איכותי, קוהרנטי ויצירתי להמשך תהליך כתיבת הסיפורים. המטרה הייתה להכריע איזה מבין שני האלגוריתמים Plotto או Plot Genie מתאים יותר לצרכי הפרויקט מבחינת איכות שלדי הסיפור.

**אופן ביצוע הניסוי**

* הופקו 50 שלדי סיפורים שונים על ידי כל אחד מהאלגוריתמים .
* כל שלד הורחב לסיפור מלא בעזרת בינה מלאכותית, ולאחר מכן בוצעה ביקורת ספרותית אוטומטית לכל אחד מהסיפורים.
* הציונים חושבו באופן משוקלל על פי מדדים מוגדרים מראש (דמויות, קונפליקט מלאכת כתיבה והיגיון). לאחר איסוף הנתונים, בוצעה השוואה סטטיסטית בין תוצאות שני האלגוריתמים.
* כל קטגוריה הינה משוקללת לפי חישוב מחדש של היחס לפי המסמך שקיבלנו.
* כל שאלה בכל מדד משוקללת באופן שונה בהתאם להשפעתה על טיב הסיפור.
* טווח ציוני השאלות נאים בין 0 – 4, כאשר 0 – לא רלוונטי, 1 – חוסר הסכמה,

2 – הסכמה חלקית, 3 – הסכמה מרובה ו- 4 – הסכמה מוחלטת.

* כנספח לניסוי זה מצורפת טבלה ובה ציון כל סיפור עבור אחד מהמדדים והציון הכולל שקיבל הסיפור + הערכים הסטטיסטיים. ציוני המדדים בטבלה הם לאחר התחשבות במשקלן מכלל הציון הסופי, כלומר אם ציון מדד הדמויות בסיפור הוא 1.155 הציון שקיבל בפועל מתוך 4 נקודות הוא כיוון שמשקל הקטגוריה הוא 30% מהציון הכולל.

**מהלך ההשוואה:**

בוצע ניתוח של**50 סיפורים שונים** שנוצרו על ידי כל אחד מהאלגוריתמים – Plotto ו־Plot Genie (ללא התאמה בין הזוגות).  
כל סיפור נוצר על בסיס שלד שהופק מהאלגוריתם, הורחב באמצעות מודל שפה, ולאחר מכן עבר **ביקורת ספרותית אוטומטית**.  
הביקורת התבצעה באמצעות מערכת המבוססת על שאלות שנגזרו ממאמר להערכת תסריטים (מצורף בנספח א') וכללה את המדדים: **דמויות**, **קונפליקט**, **מלאכת כתיבה**, ו־**היגיון עלילתי**.

הציונים חושבו בצורה משוקללת לפי הגדרת משקל שונה לכל שאלה בכל מדד כאשר קיים ציון שהתקבל עבור כל מדד וציון כולל המורכב מהם.  
  
המבחנים המתוארים בוצעו עבור ציון הכולל וציונם של כל אחד מהמדדים.

תחילה בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

לאחר מכן ביצענו בדיקה האם ההנחות של מבחן t לא-מזווג אכן מתקיימות.   
ביצענו מבחן Levene לבדיקת שוויון שונות ומבחן Shapiro-Wilk לבדיקה האם הנתונים מתפלגים נורמלית.

לבסוף ביצענו מבחן – מנ־וויטני (Mann-Whitney U test).

כלומר, ההשוואה הסטטיסטית בוצעה באמצעות מבחן t לשתי קבוצות בלתי תלויות (**Independent t-test**) ומבחן מנ־וויטני (**Mann-Whitney U tes**t).

**ניתוח ומסקנות:**

נספח א' מכיל את הטבלה המכיל את כלל תוצאות הניסוי.

**ניתוח עבור ציון כולל**

|  |  |  |
| --- | --- | --- |
| מדד סטטיסטי | Plotto | Plot Genie |
| ממוצע | 3.64 | 3.63 |
| חציון | 3.68 | 3.64 |
| סטיית תקן | 0.14 | 0.11 |

**שלב א' –**

בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): אין הבדל בין ממוצע הציונים של Plotto לבין Plot Genie.

H₁ (השערה חלופית): קיים הבדל בין ממוצע הציונים של Plotto לבין Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, אין הבדל מובהק בין ממוצעי הציונים של שתי השיטות, כלומר אין שיטה עדיפה בצורה מובהקת.**

**שלב ב' –**

נבצע בדיקת לקיום ההנחות של מבחן t לא מזווג.

נבצע מבחן Levene – לבדיקת שוויון שונות

H₀ (השערת האפס): השונות בשתי הקבוצות שווה.

H₁ (השערה חלופית): השונות בשתי הקבוצות שונה.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, השונות בשתי הקבוצות שווה.**

נבצע מבחן Shapiro-Wilk – לבדיקת נורמליות (האם הנתונים מתפלגים נורמלית)

H₀ (השערת האפס): הנתונים בקבוצה מתפלגים נורמלית.

H₁ (השערה חלופית): הנתונים אינם מתפלגים נורמלית.

תוצאה:

עבור Plotto:

עבור Plot Genie:

מסקנה:

כיוון ש־p < 0.05, הנתונים אינם מתפלגים נורמלית.

**דוחים את H0 – כלומר, התנאי של נורמליות לא מתקיים, לכן לא ניתן להסתמך על מבחן t, ונדרש לעבור למבחן לא פרמטרי – מנ-וויטני.**

**שלב ג' –**

בוצע מבחן מנ-וויטני (Mann-Whitney U) כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): ההתפלגות של הציונים ב־Plotto זהה לזו של Plot Genie.

H₁ (השערה חלופית): ההתפלגות של הציונים שונה בין Plotto ל־Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, ההתפלגות של הציונים ב־Plotto זהה לזו של Plot Genie, כלומר אין שיטה עדיפה בצורה מובהקת.**

**מסקנות עבור ניתוח ציון כולל:**

בכל אחד מהמבחנים, השערת האפס התקבלה – אין הבדל מובהק בין Plotto ל־Plot Genie במדד הציון הכולל.

**ניתוח עבור קריטריון - דמויות**

|  |  |  |
| --- | --- | --- |
| מדד סטטיסטי | Plotto | Plot Genie |
| ממוצע | 1.105 | 1.115 |
| חציון | 1.11 | 1.125 |
| סטיית תקן | 0.05 | 0.041 |

**שלב א' –**

בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): אין הבדל בין ממוצע ציוני הדמויות של Plotto לבין Plot Genie.

H₁ (השערה חלופית): קיים הבדל בין ממוצע ציוני הדמויות של Plotto לבין Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, אין הבדל מובהק בין ממוצעי ציוני הדמויות של שתי השיטות, כלומר אין שיטה עדיפה בצורה מובהקת.**

**שלב ב' –**

נבצע בדיקת לקיום ההנחות של מבחן t לא מזווג.

נבצע מבחן Levene – לבדיקת שוויון שונות

H₀ (השערת האפס): השונות בשתי הקבוצות שווה.

H₁ (השערה חלופית): השונות בשתי הקבוצות שונה.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, השונות בשתי הקבוצות שווה.**

נבצע מבחן Shapiro-Wilk – לבדיקת נורמליות (האם הנתונים מתפלגים נורמלית)

H₀ (השערת האפס): הנתונים בקבוצה מתפלגים נורמלית.

H₁ (השערה חלופית): הנתונים אינם מתפלגים נורמלית.

תוצאה:

עבור Plotto:

עבור Plot Genie:

מסקנה:

כיוון ש־p < 0.05, הנתונים אינם מתפלגים נורמלית.

**דוחים את H0 – כלומר, התנאי של נורמליות לא מתקיים, לכן לא ניתן להסתמך על מבחן t, ונדרש לעבור למבחן לא פרמטרי – מנ-וויטני.**

**שלב ג' –**

בוצע מבחן מנ-וויטני (Mann-Whitney U) כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): ההתפלגות של ציוני הדמויות ב־Plotto זהה לזו של Plot Genie.

H₁ (השערה חלופית): ההתפלגות של ציוני הדמויות שונה בין Plotto ל־Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, ההתפלגות של ציוני הדמויות ב־Plotto זהה לזו של Plot Genie, כלומר אין שיטה עדיפה בצורה מובהקת.**

**מסקנות עבור ניתוח מדד דמויות:**

בכל אחד מהמבחנים, השערת האפס התקבלה – אין הבדל מובהק בין Plotto ל־Plot Genie במדד הדמויות.

**ניתוח עבור קריטריון - קונפליקט**

|  |  |  |
| --- | --- | --- |
| מדד סטטיסטי | Plotto | Plot Genie |
| ממוצע | 1.145 | 1.13 |
| חציון | 1.155 | 1.14 |
| סטיית תקן | 0.046 | 0.036 |

**שלב א' –**

בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): אין הבדל בין ממוצע ציוני הקונפליקט של Plotto לבין Plot Genie.

H₁ (השערה חלופית): קיים הבדל בין ממוצע ציוני הקונפליקט של Plotto לבין Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, אין הבדל מובהק בין ממוצעי ציוני הקונפליקט של שתי השיטות, כלומר אין שיטה עדיפה בצורה מובהקת.**

**שלב ב' –**

נבצע בדיקת לקיום ההנחות של מבחן t לא מזווג.

נבצע מבחן Levene – לבדיקת שוויון שונות

H₀ (השערת האפס): השונות בשתי הקבוצות שווה.

H₁ (השערה חלופית): השונות בשתי הקבוצות שונה.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, השונות בשתי הקבוצות שווה.**

נבצע מבחן Shapiro-Wilk – לבדיקת נורמליות (האם הנתונים מתפלגים נורמלית)

H₀ (השערת האפס): הנתונים בקבוצה מתפלגים נורמלית.

H₁ (השערה חלופית): הנתונים אינם מתפלגים נורמלית.

תוצאה:

עבור Plotto:

עבור Plot Genie:

מסקנה:

כיוון ש־p < 0.05, הנתונים אינם מתפלגים נורמלית.

**דוחים את H0 – כלומר, התנאי של נורמליות לא מתקיים, לכן לא ניתן להסתמך על מבחן t, ונדרש לעבור למבחן לא פרמטרי – מנ-וויטני.**

**שלב ג' –**

בוצע מבחן מנ-וויטני (Mann-Whitney U) כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): ההתפלגות של ציוני הדמויות ב־Plotto זהה לזו של Plot Genie.

H₁ (השערה חלופית): ההתפלגות של ציוני הדמויות שונה בין Plotto ל־Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p < 0.05, נדחה את השערת האפס.

**דוחים את H₀ – כלומר, ההתפלגות של ציוני הדמויות ב־Plotto שונה מזו של Plot Genie. חושבו דירוגי כלל הסיפורים בשתי הקבוצות ונמצא כי ממוצע הדירוג של Plotto גבוה משמעותית מזה של Plot Genie (56.74 לעומת 44.26 בהתאמה). ממצא זה מצביע על כך שסיפורי Plotto מדורגים גבוה יותר באופן עקבי במדד קונפליקט, ומכאן שניתן להסיק כי Plotto עדיף סטטיסטית בקריטריון זה.**

**מסקנות עבור ניתוח מדד קונפליקט:**

בכל אחד מהמבחנים, השערת האפס התקבלה, מלבד מבחן – קיים הבדל מובהק בין Plotto ל־Plot Genie במדד הקונפליקט – ו־Plotto עדיף סטטיסטית על פני Plot Genie לפי דירוגי הציונים שהוזנו.

**ניתוח עבור קריטריון – מלאכת כתיבה**

|  |  |  |
| --- | --- | --- |
| מדד סטטיסטי | Plotto | Plot Genie |
| ממוצע | 0.898 | 0.896 |
| חציון | 0.9125 | 0.9125 |
| סטיית תקן | 0.04 | 0.04 |

**שלב א' –**

בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): אין הבדל בין ממוצע ציוני מלאכת הכתיבה של Plotto לבין Plot Genie.

H₁ (השערה חלופית): קיים הבדל בין ממוצע ציוני מלאכת הכתיבה של Plotto לבין Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, אין הבדל מובהק בין ממוצעי ציוני מלאכת הכתיבה של שתי השיטות, כלומר אין שיטה עדיפה בצורה מובהקת.**

**שלב ב' –**

נבצע בדיקת לקיום ההנחות של מבחן t לא מזווג.

נבצע מבחן Levene – לבדיקת שוויון שונות

H₀ (השערת האפס): השונות בשתי הקבוצות שווה.

H₁ (השערה חלופית): השונות בשתי הקבוצות שונה.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, השונות בשתי הקבוצות שווה.**

נבצע מבחן Shapiro-Wilk – לבדיקת נורמליות (האם הנתונים מתפלגים נורמלית)

H₀ (השערת האפס): הנתונים בקבוצה מתפלגים נורמלית.

H₁ (השערה חלופית): הנתונים אינם מתפלגים נורמלית.

תוצאה:

עבור Plotto:

עבור Plot Genie:

מסקנה:

כיוון ש־p < 0.05, הנתונים אינם מתפלגים נורמלית.

**דוחים את H0 – כלומר, התנאי של נורמליות לא מתקיים, לכן לא ניתן להסתמך על מבחן t, ונדרש לעבור למבחן לא פרמטרי – מנ-וויטני.**

**שלב ג' –**

בוצע מבחן מנ-וויטני (Mann-Whitney U) כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): ההתפלגות של ציוני מלאכת הכתיבה ב־Plotto זהה לזו של Plot Genie.

H₁ (השערה חלופית): ההתפלגות של ציוני מלאכת הכתיבה שונה בין Plotto ל־Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, ההתפלגות של ציוני מלאכת הכתיבה ב־Plotto זהה לזו של Plot Genie, כלומר אין שיטה עדיפה בצורה מובהקת.**

**מסקנות עבור ניתוח מדד מלאכת הכתיבה:**

בכל אחד מהמבחנים, השערת האפס התקבלה – אין הבדל מובהק בין Plotto ל־Plot Genie במדד מלאכת הכתיבה.

**ניתוח עבור קריטריון – היגיון**

|  |  |  |
| --- | --- | --- |
| מדד סטטיסטי | Plotto | Plot Genie |
| ממוצע | 0.495 | 0.487 |
| חציון | 0.495 | 0.495 |
| סטיית תקן | 0.044 | 0.037 |

**שלב א' –**

בוצע מבחן t לא-מזווג כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): אין הבדל בין ממוצע ציוני ההיגיון של Plotto לבין Plot Genie.

H₁ (השערה חלופית): קיים הבדל בין ממוצע ציוני ההיגיון של Plotto לבין Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, אין הבדל מובהק בין ממוצעי ציוני ההיגיון של שתי השיטות, כלומר אין שיטה עדיפה בצורה מובהקת.**

**שלב ב' –**

נבצע בדיקת לקיום ההנחות של מבחן t לא מזווג.

נבצע מבחן Levene – לבדיקת שוויון שונות

H₀ (השערת האפס): השונות בשתי הקבוצות שווה.

H₁ (השערה חלופית): השונות בשתי הקבוצות שונה.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, השונות בשתי הקבוצות שווה.**

נבצע מבחן Shapiro-Wilk – לבדיקת נורמליות (האם הנתונים מתפלגים נורמלית)

H₀ (השערת האפס): הנתונים בקבוצה מתפלגים נורמלית.

H₁ (השערה חלופית): הנתונים אינם מתפלגים נורמלית.

תוצאה:

עבור Plotto:

עבור Plot Genie:

מסקנה:

כיוון ש־p < 0.05, הנתונים אינם מתפלגים נורמלית.

**דוחים את H0 – כלומר, התנאי של נורמליות לא מתקיים, לכן לא ניתן להסתמך על מבחן t, ונדרש לעבור למבחן לא פרמטרי – מנ-וויטני.**

**שלב ג' –**

בוצע מבחן מנ-וויטני (Mann-Whitney U) כדי לבדוק האם ההבדלים בין ממוצעי שתי הקבוצות מובהקים סטטיסטית.

H₀ (השערת האפס): ההתפלגות של ציוני ההיגיון ב־Plotto זהה לזו של Plot Genie.

H₁ (השערה חלופית): ההתפלגות של ציוני ההיגיון שונה בין Plotto ל־Plot Genie.

תוצאה:

מסקנה:

כיוון ש־p > 0.05, נקבל את השערת האפס.

**מקבלים את H₀ – כלומר, ההתפלגות של ציוני ההיגיון ב־Plotto זהה לזו של Plot Genie, כלומר אין שיטה עדיפה בצורה מובהקת.**

**מסקנות עבור ניתוח מדד ההיגיון:**

בכל אחד מהמבחנים, השערת האפס התקבלה – אין הבדל מובהק בין Plotto ל־Plot Genie במדד ההיגיון.

**מסקנות מהניסוי**

מטרת הניסוי הייתה לבחון האם קיימים הבדלים מובהקים בין שני אלגוריתמים לבניית שלד עלילה **Plotto** **וPlot Genie** בארבעה מדדי הערכה (והציון הכולל שלהם): **דמויות, קונפליקט, מלאכת כתיבה והיגיון.**

להלן סיכום הממצאים העיקריים:

* **ציון כולל:**  
  לא נמצאה עדות להבדל מובהק בין שני האלגוריתמים. גם מבחן t וגם מבחן מנ-וויטני הצביעו על קבלת השערת האפס, כלומר – איכות הסיפורים בשני האלגוריתמים דומה במדד הכולל.
* **מדד דמויות:**  
  לא נמצאה עדות להבדל מובהק בין שני האלגוריתמים. גם מבחן t וגם מבחן מנ-וויטני הצביעו על קבלת השערת האפס, כלומר – איכות הסיפורים בשני האלגוריתמים דומה במדד הדמויות.
* **מדד קונפליקט:**  
  נמצא הבדל מובהק בין שני האלגוריתמים. מבחן מנ-וויטני הצביע על דחיית השערת האפס, כאשר ציוני Plotto היו גבוהים באופן מובהק מ Plot Genie. משמעות הדבר היא ש־**Plotto מייצר שלדים עם קונפליקטים איכותיים יותר**.
* **מדד מלאכת הכתיבה:**  
  לא נמצאה עדות להבדל מובהק בין שני האלגוריתמים. גם מבחן t וגם מבחן מנ-וויטני הצביעו על קבלת השערת האפס, כלומר – איכות הסיפורים בשני האלגוריתמים דומה במדד מלאכת הכתיבה.
* **מדד היגיון:**  
  לא נמצאה עדות להבדל מובהק בין שני האלגוריתמים. גם מבחן t וגם מבחן מנ-וויטני הצביעו על קבלת השערת האפס, כלומר – איכות הסיפורים בשני האלגוריתמים דומה במדד ההיגיון.

**לסיכום, הניסוי מצביע על כך ששני האלגוריתמים מפיקים שלדים באיכות דומה ברוב הקריטריונים, למעט מדד הקונפליקט, שבו Plotto נמצא כעדיף מובהקות. נתון זה עשוי להצביע על יתרון יחסי של Plotto ביצירת קונפליקטים עלילתיים חזקים ומשכנעים יותר – מרכיב מרכזי בכתיבת סיפורים.**

**בנוסף לדו"ח זה, מצ"ב תיקיות הכוללת את כל הסיפורים שהופקו + דוחות ביקרות טיב הסיפור:**

1. Plotto – כולל את כל הסיפורים ודוחות הביקורת עבורן עבור סיפור שנוצרו באמצעות שלד שנוצר ע"י אלגוריתם Plotto.
2. Plot Genie– כולל את כל הסיפורים ודוחות הביקורת עבורן עבור סיפור שנוצרו באמצעות שלד שנוצר ע"י אלגוריתם Plot Genie.