**בינה עסקית – מטלה 1**

שאלות מחקר:

Unsupervised:

1. האם ניתן לחזות מי המדינה שתנצח במונדיאל הבא?

Supervised:

1. האם יש קשר בין המדינה שהבקיעה יותר גולים עד המחצית למדינה המנצחת?

KPI's

1. - **ההפרש בין מספר הגולים במשחק של הקבוצה המפסידה לקבוצה המנצחת לא יעלה מעל גול אחד.**

S- המטרה מוגדרת היטב ובצורה ספציפית.

M- ניתן למדוד את ההפרש בין הגולים בכל משחק בין 2 הקבוצות.

1. הדבר בר השגה, מטרה ישימה בחישוב פשוט.

T- רלבנטי למטרות ארוכות טווח.

R- הגדרנו כי הKPI ימדד בכל משחק מהתחילתו ועד סופו – 90 דק (כמעט תמיד).

* **כמות ההפסדים של כל מדינה יירד ב25% עד למונדיאל הבא.**

S- השאלה האם קבוצה הפסידה, מקבלת תשובה חד משמעית וברורה.

M- ניתן למדוד ולחשב את אחוז ההפסדים המשתנה ממונדיאל למונדיאל.

1. הדבר בר השגה, באמצעות אימונים רבים והכנה מספיק טובה למונדיאל.

R- היעד רלבנטי עבור המטרה – לזכות במונדיאל.

T- עד למונדיאל הבא – 4 שנים.

* **עד המחצית הראשונה, יש להבקיע לפחות גול אחד ולהיות ביתרון.**

S- הבקעת גול אחד, הוא דבר המוגדר היטב וברור מאוד האם הדבר אכן קרה.

M- הבקעת גול אחד, משהו מדיד צריך לקבל את מספר 1.

1. בר השגה, באמצעות אימונים טובים והתקפה חזקה.

R- רלבנטי למטרה, עוזר ביתרון בשלב מוקדם למשחק.

T- מחצית אחת – 45 דק.

* **כמות הנצחונות של כל מדינה יגדל ב20% עד המונדיאל הבא.**

S- הכמות הנדרשת מוגדרת באופן ספציפי וברור.

M- ניתן למדוד את כמות הנצחונות באמצעות חישוב פשוט של הנתונים.

1. הדבר בר השגה, באמצעות אימונים רבים והכנה מספיק טובה למונדיאל.

R- רלבנטי למטרות ארוכות טווח.

T- עד למונדיאל הבא – 4 שנים.

* **הובקע לפחות גול אחד במחצית הראשונה. (המחצית הראשונה לא הסתיימה בתיקו)**

S- הבקעת גול אחד, הוא דבר המוגדר היטב וברור מאוד האם הדבר אכן קרה.

M- הבקעת גול אחד, משהו מדיד צריך לקבל את מספר 1.

1. בר השגה, באמצעות אימונים טובים והתקפה חזקה.

R- רלבנטי למטרה, עוזר ביתרון בשלב מוקדם למשחק.

T- מחצית אחת – 45 דק.

* הקבוצה המובילה במחצית הראשונה תהיה המנצחת במשחק כולו

S- הקבוצה המובילה מוגדרת באופן חד משמעי (בהתאם לקבוצה בעלת מספר הגולים הגבוה יותר)

M- ניתן לדעת מי הקבוצה המנצחת בהתאם לכמות הגולים, שניתנים לספירה

1. בר השגה , ניצחון במשחק הוא דבר אפשרי

R- רלבנטי למטרה

T – הזמנים מוגדרים היטב- מחצית ראשונה לעומת המשחק כולו.

**מסקנות:**

הערך של ה- info gain מעיד על כך שהקשר בין כמות הגולים עד המחצית לבין הקבוצה המנצחת הוא מועמד טוב להיות בשורש העץ, ולכן ניתן להסיק כי כמות הגולים הוא מאפיין משמעותי לחיזוי.

קישור לקוד ששימש אותנו לניתוח הנתונים: <https://colab.research.google.com/drive/1HhfldJItKyw63GLGAexY5_EQirxCG8_x?usp=sharing>

בנוסף, מצורף גם קובץ אקסל עם הraw data והניתוחים הפשוטים (ממוצע, קורלציה וכו').