**דו"ח סיכום – פרויקט סוף – בינה מלאכותית**

**סעדיה ביררוב – ת.ז 206812448**

1. **שאלת מחקר:**

בהינתן נתונים אודות אופן ההכנה, פעולות מקדימות, רקע חינוכי של הורי התלמיד ועוד לפני ביצוע מבחנים – האם ניתן לחזות את הציונים שיתקבלו במבחנים בנושאים:

* מתמטיקה
* כתיבה
* קריאה

ניתוח וחיזוי הנתונים מתבצע באמצעות ספריית brain.js אשר מאפשרת ליצור "רשת נוירונים" , להגדירה באמצעות תכונות כמו מספר איטרציות ללימוד, דרישת סף למרווח שגיאה בחיזוי, מודל נתון מראש בסיסי אשר מגדיר לרשת את אופן עבודתה וכלים נוספים המשמשים לבדיקה של ביצועי הרשת במהלך קומפילציה.

לימוד המכונה מתבצע על ידי אימון שלה עם נתונים מלאים הכוללים את הנתונים הנ"ל והציונים אשר הופקו מתנאים כאלה, החיזוי מתבצע על ידי הזנה של הנתונים ללא הציונים וקבלת ציון חזוי לתלמיד אשר מבצע את המבחנים תחת תנאים אלה.

1. הנתונים הופקו מהאתר <https://www.kaggle.com/datasets/bhavikjikadara/student-study-performance> , אשר מספק נתונים מוסדרים חלקית בצורה של גליון אקסל (CSV).

נתונים אלה הומרו לפורמט json ואוחסנו בדף .js בתיקיית הפרויקט. **לצורך האימון והשימוש ברשת הנוירונים הנתונים נורמלו למספרים באופן הבא:**

**Male / female** – נורמלו למספרים 0 ו 1 בהתאמה. (מגדר התלמיד)

**Group A – E** - 1-5 בהתאמה. (קבוצת גזע)

**רמת השכלה של הורי התלמיד :**

High school - 0

Associate degree - 1

Master’s degree- 2

Some college - 3

Bachelor’s degree – 4

**קבלת ארוחה לפני המבחן:**

Standard - 1

Free / reduced - 0

**הכנה רלוונטית לבחינה לפני ביצועה:**

Complete / not complete: bool – תנאי בוליאני (true \ false).

רשת נוירונים LSTM (Long Short Term Memory) נבנתה ומקבלת נתוני קלט של משפטים יום יומיים ומטרתה היא לסווג משפטי קלט שהיא מקבלת לנושאים הכלליים שעליה אומנה. (רשת זו נמצאת בקוד אך כשלה בביצוע).

1. **הגדרת רשת הנוירונים:**

יצירת מופע של רשת נוירונים עם ההגדרות הבאות:

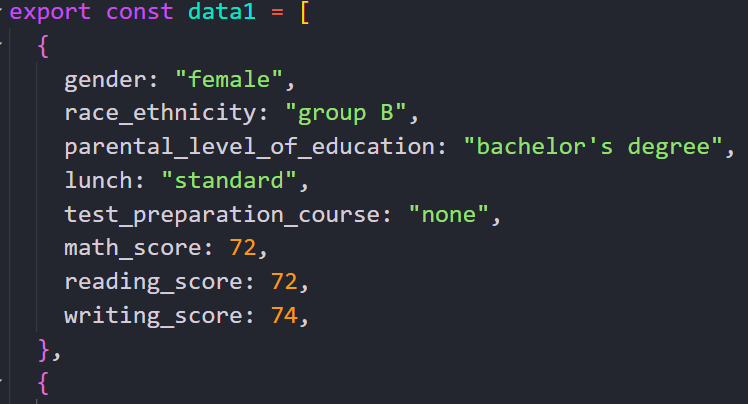
* 5 נתוני קלט (נתוני תלמיד)
* 3 נתוני פלט (ציונים)
* 3 שכבות נסתרות של 20 נוירונים לכל אחת.
* אקטוב בשיטת “sigmoid”.
* מרווח שגיאה של 0.005
* 20,000 איטרציות ללימוד

**הגדרת רשת LSTM ( Long Short Term Memory):**

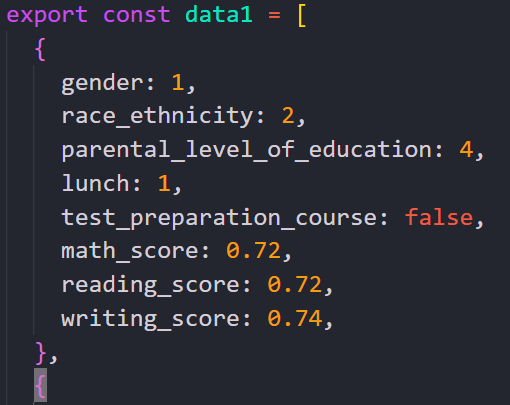
* 5000 איטרציות ללימוד
* מרווח שגיאה של 0.011

1. **תהליך האימון :**

ברשת הנוירונים לחיזוי ציוני מבחנים על סמך נתוני תלמידים – המידע שהומר לקובץ JSON באופן הבא:



בעזרת הפונציה CreateTrainingData המידע הומר לנתונים מספריים, לאחר פעולת הפונקציה המידע מוצג באופן הבא:



קובץ הנתונים מכיל 6000 מופעים של דוגמאות אלה, כמות זו פוצלה ל 3 מחרוזות נפרדות, בהרצת המודל המכונה "מתאמנת" כל פעם על מחרוזת אחת.

לאחר הרצת האימון, מוזנים למכונה נתוני הקלט (5) ולאחר החיזוי, היא מחזירה 3 נתוני פלט שהם הציונים של הסטודנטים.

1. **תוצאות הניסויים:**

תוצאות הניסויים הניבו ערך בין 0 ל 1 אשר מסמל את הציון הרלוונטי בכל נושא, בהכפלת המספר והמרתו למספר שלם התקבל הציון שהמכונה חזתה.

**ברשת LSTM** – התוצאות היו מעורבות, בסיווג של נושאים כלליים תחילה התקבלו רצפים של אותיות ללא הגיון או משמעות, לאחר שינויים נוספים בהגדרת הרשת התקבלו רצפי אותיות ללא הגיון ובנוסף מילת הסיווג הרצויה (באחוזי דיוק נמוכים).

1. לאחר ניתוח תוצאות ניתן לראות שבדומה לקובץ המידע שניתן למכונה לאימון, אין קורלציה חד משמעית בין התנאים שבהם מצויים התלמידים לציוני הבחינות שלהם, רקע אקדמי של הורי התלמידים אכן תורם לציונים גבוהים יותר אך שאר הנתונים לא בהכרח משפיעים. הסקת מסקנה זו מתבססת על התחזיות שהמכונה סיפקה – כמובן בהתחשב בעובדה שסביר להניח שהמכונה לא הוגדרה בצורה האידיאלית למשימה ושמדובר בספריה הבנויה על גאווה סקריפט.
2. ביצועי הרשת והיכולת שלה ללמוד מהדאטה שסופקה מסתמכת רבות על החומרה של המכונה שבה משתמשים להרצה, אי דיוק של התוצאות נובע מהגדרה לא נכונה של הרשת או חומרה לא מספקת; כמות איטרציות, שכבות נסתרות של נוירונים, חוזק הרשת האינטרנטית וזיכרון ראם פנוי.
3. **ניתן לסכם את הפרויקט ותוצאותיו בנקודות הבאות:**

* לצורך הרצה של רשת מתפקדת ויעילה, יש להעביר למכונה כמות מידע רבה בשלב האימון, שלב זה בתהליך ההרצה לוקח זמן ודורש משאבים רבים מהחומרה של המכונה.
* בניתוח התוצאות החזויות שהתקבלו מהמכונה ניתן לראות שהיא מחלקת ציונים מעט נמוכים מהצפוי אך בטווח מקובל והגיוני.