דו"ח מסכם: חיזוי מחירי רכב וזיהוי יצרן על פי שם דגם

8.1.1 ניסוח שאלת המחקר:

שאלת המחקר המרכזית בפרויקט זה היא:  
 האם ניתן לחזות את מחיר המכירה של רכב משומש בהתבסס על תכונותיו, כגון יצרן, דגם, שנת ייצור  
 מצב ונסועה, באמצעות רשתות LSTM?

כמו כן, נחקרה שאלה משנית: האם ניתן לזהות את יצרן הרכב בהתבסס על שם הדגם בלבד?

8.1.2 תיאור איסוף הנתונים והמרתם

מקור נתונים:  
 הנתונים נאספו מקובץ CSV המכיל מידע על מחירי מכירה של רכבים משומשים בקליפורניה.  
 https://www.kaggle.com/

תכונות: הנתונים כוללים תכונות כגון יצרן, דגם, שנת ייצור, מצב, קילומטרז, סוג תיבת הילוכים,  
 מחיר ועוד.

עיבוד נתונים:  
נתונים חסרים או שגויים הוסרו.  
ערכים מספריים, כגון שנת ייצור, מצב ונסועה, נורמלו לטווח בין 0 ל-1.  
הנתונים חולקו לערכות אימון, ולידציה ובדיקה.

8.1.3 ארכיטקטורת רשת הנוירונים ורשת ה-LSTM

רשת LSTM לחיזוי מחיר:  
רשת LSTMTimeStep מ-Brain.js.  
שכבות נסתרות: 2 שכבות עם 64 נוירונים בכל שכבה.  
קלט: שנת ייצור, מצב וקילומטרז (לאחר נרמול).  
פלט: מחיר (לאחר דנרמול).

רשת LSTM לזיהוי יצרן:  
 רשת LSTM מ-Brain.js.  
 שכבה נסתרת אחת עם 32 נוירונים.  
 קלט: שם הדגם (כמחרוזת).  
 פלט: שם היצרן (כמחרוזת).

8.1.4 תהליך האימון

רשת חיזוי מחיר:  
הרשת אומנה על ערכת האימון תוך שימוש בפרמטרים הבאים:  
מספר איטרציות: 50.  
סף שגיאה: 0.025.

רשת זיהוי יצרן:  
הרשת אומנה על ערכת האימון תוך שימוש בפרמטרים הבאים:  
מספר איטרציות: 60.  
סף שגיאה: 0.005.

8.1.5 תוצאות הניסויים

רשת חיזוי מחיר:  
התוצאות הראשוניות מראות כי הרשת מסוגלת ללמוד קשר בין תכונות הרכב למחירו.  
יש צורך בניסויים נוספים והערכה מדויקת יותר של ביצועי הרשת, עם התחשבות במדינה באה נמכר הרכב, גודלו, נפח מנוע, סוג תיבת הילוכים ועוד.

רשת זיהוי יצרן:  
הרשת הצליחה ללמוד קשר בין שמות דגמים ליצרנים.  
יש צורך בהרחבת נתוני נאימון

8.1.6 ניתוח תוצאות והסקת מסקנות

רשת חיזוי מחיר:

רשתות LSTM יכולות לשמש לחיזוי מחירי רכב, אך ביצועיהן תלויים בגורמים כמו כמות הנתונים, איכות הנתונים וארכיטקטורת הרשת.

יש לבחון שילוב של תכונות נוספות, כמו דגם הרכב, סוג תיבת הילוכים, צבע וכו',  
על מנת לשפר את הדיוק.

רשת זיהוי יצרן:

רשתות LSTM יכולות לשמש לזיהוי יצרן הרכב בהתבסס על שם הדגם, אך יש לבחון את ביצועיהן על   
מגוון רחב יותר של דגמים ויצרנים.

8.1.7 דיון בביצועי הרשת ובמגבלותיה

רשת חיזוי מחיר:  
ייתכן שרשת LSTM עם מספר קטן של שכבות ונוירונים ומספר איטרציות נמוך אינה מסוגלת ללמוד קשרים מורכבים בין תכונות הרכב למחירו.

רשת זיהוי יצרן:  
ייתכן שהרשת תלמד קשרים שגויים או תתקשה לזהות יצרנים של דגמים נדירים.

יש צורך בבדיקה של טכניקות עיבוד שפה טבעית (NLP) נוספות, כמו Word2Vec או Embeddings,  
על מנת לשפר את הייצוג של שמות דגמים. (היה נסיון להשתמש בEmbeddings מספריית tensorflow.js אך לא מואיל, לכו לא נכנס לגרסה הסופית)

8.1.8 סיכום ומסקנות

פרויקט זה הראה את הפוטנציאל של רשתות LSTM לחיזוי מחירי רכב ולזיהוי יצרנים בהתבסס  
על שמות דגמים. יש צורך במחקר ופיתוח נוספים על מנת לשפר את דיוקן ולבחון את יכולותיהן  
בתחום זה.

המלצות להמשך:  
שילוב תכונות נוספות, כמו דגם, סוג תיבת הילוכים, רמת גימור, נפח מנוע מדינה וכו', ברשת חיזוי המחיר.  
השוואת ביצועי רשת LSTM עם מודלים אחרים לחיזוי מחירים, כמו רשתות נוירונים מסוג אחר.  
בדיקת הרשתות על נתונים ממדינות או אזורים שונים.