**נאום:**

**הקדמה:**

במסגרת פרוייקט הגמר שלנו רצינו מאוד לבחור, לבצע ולפתח, רעיון בנושא משמעותי אשר יוביל לשינוי חיובי וכן יתרום ויעזור לאוכלוסייה מסוימת בחברה.

לאחר חשיבה וחקירה בנושא הגענו למסקנה כי קיים בחברה קושי רב באינטראקציה בין אנשים חירשים לשומעים.

הקושי מתבטא בכך שאנשים חירשים מדברים באמצעות שפת הסימנים, ואילו רוב האנשים השומעים אינם מבינים כלל וכלל את השפת הסימנים, ובכך נוצר קוצר בתקשורת ואין באפשרותם דרך לדבר ולהבין זה את זה בצורה טובה.

אנשים חירשים ואילמים משתמשים בשפת הסימנים של תנועות ידיים כדי לתקשר, ומכאן שאנשים רגילים נתקלים בבעיות בזיהוי שפתם על ידי סימנים שנעשו. מכאן שיש צורך במערכות המזהות את הסימנים השונים ומעבירות את המידע לאנשים שלא בקיאים בשפה.

שפת סימנים היא כינוי לכל אחת מהשפות הטבעיות שמשתמשות בתנועות ידיים, מחוות גוף והבעות פנים כאמצעים העיקריים להעברת מסרים וליצירת תקשורת בין הדוברים. האוכלוסייה העיקרית שמשתמשת בשפת סימנים היא אוכלוסיית החירשים וכבדי-השמיעה, בעיקר אלה שנולדו כבדי-שמיעה או שלקו בשמיעתם בגיל צעיר. במקומות שונים בעולם משמשות שפות סימנים שונות, ואין בהכרח יכולת לדובר שפת סימנים אחת להבין דובר בשפת סימנים אחרת.

לכן, החלטנו במסגרת הפרויקט שלנו בהנדסת תוכנה לבנות אפליקציה שמתרגמת בזמן אמת את שפת הסימנים.

האפליקציה שבנינו מתרגמת באמצעות המצלמה של הטלפון את אותיות הא"ב באנגלית משפת הסימנים לאותיות באנגלית.

בישראל יש כשבע מאות אלף כבדי שמיעה וחירשים (כ-8% מן האוכלוסייה), מתוכם כעשרים אלף איש משתמשים בשפת הסימנים הישראלית במידה זו או אחרת, והיא משמשת שפת אם לחלק ניכר מהם, בפרט הצעירים שבהם שהתחנכו במערכת החינוך הישראלית. שפת הסימנים הישראלית מוכרת בישראל כשפה המשמשת במערכת החינוך לחירשים, וכשפת התקשורת בין חירשים לשומעים.

בעקבות הפרויקט שבנינו בהנדסת תוכנה, גילינו שמאגר המידע של שפת הסימנים בעברית מצומצם מאוד. בעקבות בעיה זו חשבנו איך זה אפשרי שבשנת 2022 אין מספיק מידע מתויג המאפשר לבנות מודל איכותי במגוון שפות למעט השפה האנגלית?

ולכן במסגרת פרויקט הגמר החלטנו לבנות מודל לתרגום שפת הסימנים בשפה העברית.

בעזרת המנחה שלנו גיל בן ארצי, החלטנו לנסות לקדם פתרון לבעיה זאת.

**הבעיה:**

שפת סימנים אינה תלויה בשפה המדוברת ואינה שואלת ממנה בהכרח את אוצר המילים ואת כללי הדקדוק והתחביר. כלומר, היא אינה תרגום מילה-במילה של השפה המדוברת (בניגוד ל"שפות המסומנות").

דוגמה בולטת לכך היא שפת הסימנים הבריטית ושפת הסימנים האמריקאית, שהן שפות שונות מאוד והדובר בשפה אחת לא יבין את הדובר בשפה השנייה - זאת למרות העובדה שגם בבריטניה וגם בארצות הברית מדברים באותה שפה, אנגלית.

עובדה זאת גורמת למחסור במאגרי מידע גדולים ואיכותיים של מידע מתויג בכל הקשור לשפת הסימנים בשפות שונות. כיום לא ניתן למצוא מגוון רב של מודלים איכותיים לזיהוי שפת הסימנים בשפות שאינן בשפה האנגלית.

בעקבות חוסר המידע המתויג הבנו שלא נצליח לקדם את הבעיה באמצעות אלגוריתמים בשיטת **supervised**.

למודל בשיטת supervised נדרש מידע מתויג רב מכל label ובגלל שמדובר בשפה אז יש מספר רב של מילים, ולכן החלטנו לחקור ולנסות אלגוריתמים שהם **unsupervised**.

**הפתרון:**

המטרה- תיוג המוני של דאטה בצורת תמונות של שפת הסימנים שאינו מתויג לכל שפה (ופרט לשפה העברית). במחקר שלנו לצורך POC (proof of concept) התרכזנו באותיות האלף בית העבריות.

המטרה שלנו הייתה ביצוע **clustering**.

מה זה **clustering**? שיוך של דאטא לתתי קבוצות.

Clustering נעשה בצורת unsupervised על ידי כך שהמודל לומד לזהות דמיון בין הווקטורים המייצגים את התמונות שלנו.

אחרי שלב זה קיבלנו את הדאטא שלנו מחולק לקבוצות דומות.

כעת כל שנותר לנו הוא למצוא דרך לתייג תמונות אלו.

בדקנו כמה אלגוריתמים על מנת למצוא את הטוב ביותר למטרתנו.

השתמשנו ב K Means , ובשיטות שונות לביצוע ה feature extraction.

**Kmeans-**- אלגוריתם k-מרכזים (k-means) הוא שיטה פופולרית עבור ניתוח אשכולות (Clustering) בכריית נתונים. מטרתו לחלק את התצפיות ל-k אשכולות לפי מרכזי כובד (k-means). כל תצפית משויכת לאחד מ"מרכזי הכובד". על ידי בחירה נכונה של מרכזי כובד ניתן לאתר את הקבוצות השונות.

נדרשות תצפיות רבות על מנת להשתמש במודל ותוספת של תצפיות חדשות עשויה לחייב חישוב חוזר. מדובר באלגוריתם היוריסטי שמשתמשים בו כדי לבצע חישובים שמובילים להתכנסות לפתרון מקומי (שאינו בהכרח הפתרון הטוב ביותר).

מודל זה נוטה למצוא אשכולות בעלי מרחבי מידה הניתנים להשוואה. בעוד שמנגנון ציפייה מביא למקסום ומאפשר לאשכולות להיות בעלי צורות שונות. מודל זה לא ליניארי, ולכן הוא מאפשר להתמודד עם כל סוגי ההתפלגויות האפשריות.

זהו מודל סטטיסטי שאינו מתבסס על ידע מוקדם על הנתונים אלא רק על תצפיות בפועל. כדי להשתמש באלגוריתם זה נצטרך לקחת את הדאטה המקורי ולבצע feature extraction.

**feature extraction –** בעברית: חילוץ תכונות.

**איתור תכונות ייחודיות בנתונים בהן ניתן להשתמש כדי לבצע תחזיות.**

בתהליך הזה, בעזרת אלגוריתמים הנכתבים על ידי בני האדם, המחשב יודע אילו פרטים לחפש בקלט.

בלמידת מכונה, זיהוי תבניות ועיבוד תמונה, feature extraction מתחיל מקבוצה ראשונית של נתונים נמדדים ובונה ערכים (תכונות) שנועדו להיות אינפורמטיביים ולא מיותרים, מה שמקל על שלבי הלמידה וההכללה הבאים, ובמקרים מסוימים מוביל לפרשנויות אנושיות טובות יותר.

כאשר נתוני הקלט לאלגוריתם גדולים מכדי לעבד ויש חשד כי הם מיותרים (למשל חזרה של תמונות המוצגות כפיקסלים), אז ניתן להפוך אותם לסט מצומצם של תכונות (נקרא גם וקטור תכונה). קביעת תת-קבוצה של התכונות הראשוניות נקראת בחירת תכונה.

התכונות הנבחרות צפויות להכיל את המידע הרלוונטי מנתוני הקלט, כך שניתן לבצע את המשימה הרצויה באמצעות ייצוג מופחת זה במקום הנתונים הראשוניים המלאים.

feature extraction הוא תהליך של הפחתת ממד שבאמצעותו קבוצה ראשונית של נתונים גולמיים מצטמצמת לקבוצות ניתנות לעיבוד. מאפיין של מערכי נתונים גדולים אלה הוא מספר רב של משתנים הדורשים משאבי מחשוב רבים לעיבוד. feature extraction הוא השם לשיטות שבוחרות ו/או משלבות משתנים לתכונות, תוך הפחתת כמות הנתונים שיש לעבד, ותיאור מדויק ומלא של מערך הנתונים המקורי.

התהליך של feature extraction שימושי כאשר נצטרך לצמצם את מספר המשאבים הדרושים לעיבוד מבלי לאבד מידע חשוב או רלוונטי. חילוץ תכונות יכול גם להפחית את כמות הנתונים המיותרים עבור ניתוח נתון. כמו כן, הפחתת הנתונים והמאמצים של המכונה בבניית שילובי משתנים (תכונות) מקלים על מהירות הלמידה ושלבי ההכללה בתהליך למידת המכונה.

בכך שנשתמש בסוגים שונים של feature extraction נקבל תוצאות שונות בביצוע הקלסטרינג.

הסוגים של ה feature extraction שהשתמשנו:

**VGG16-**

אלגוריתם זיהוי וסיווג אובייקטים המסוגל לסווג 1000 תמונות של 1000 קטגוריות שונות עם דיוק של 92.7%. זהו אחד האלגוריתמים הפופולריים לסיווג תמונות וקל לשימוש עם למידת העברה (transfer learning).

רשת זו מכילה סה”כ 16 שכבות מתוכם 13 [שכבות קונבלוציה](https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/) ו 3 [שכבות Dense](https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/) לצורך הסיווג. מספר הפילטרים בשכבות בקונבלוציה הולכים וגדלים  בעוד המימדים הגיאומטריים הולכים וקטנים.

החסרונות של ארכיטקטורה זו הינם:

1.שלב האימון הינו איטי יחסית.

2.התהליך מייצר מודל גדול מאוד.

**VGG19-**

VGG19 הוא גרסה של VGG אשר מורכב מ-19 שכבות (16 שכבות קונבולציה, 3 שכבות מחוברות לחלוטין, 5 שכבות MaxPool ושכבת SoftMax אחת).

מוכר גם כ-VGG19 המודל המדויק ביותר אך גם היקר יותר מבחינה חישובית.

**RestNet50-**

ResNet-50 היא רשת עצבית קונבולוציונית שעומקה 50 שכבות.

ניתן לטעון גרסה מאומנת מראש של הרשת שאומנה על יותר ממיליון תמונות ממסד הנתונים של ImageNet. הרשת המאומנת מראש יכולה לסווג תמונות ל-1000 קטגוריות אובייקטים, כגון מקלדת, עכבר, עיפרון וחיות רבות. כתוצאה מכך, הרשת למדה ייצוגי תכונות עשירים עבור מגוון רחב של תמונות. לרשת יש גודל קלט תמונה של 224 על 224.

לכל אחת מהשיטות למעלה ביצענו גם PCA:

-**PCA** Principal Component Analysis

הורדת כמות הפיצ'רים תוך שמירה על מקסימום יעילות.

PCA זוהי שיטה להפחתת ממדים המשמשת לעתים קרובות להפחתת הממדיות של מערכי נתונים גדולים, על ידי הפיכת סט גדול של משתנים לאחד קטן יותר שעדיין מכיל את רוב המידע במערך הגדול.

צמצום מספר המשתנים של מערך נתונים בא באופן טבעי על חשבון הדיוק, אבל החוכמה בהפחתת הממדיות היא להחליף מעט דיוק למען הפשטות. מכיוון שקל יותר לחקור ולהמחיש מערכי נתונים קטנים יותר והופכים את ניתוח הנתונים להרבה יותר קל ומהיר עבור אלגוריתמי למידת מכונה ללא משתנים זרים לעיבוד.

אז לסיכום, הרעיון של PCA הוא פשוט - צמצם את מספר המשתנים של מערך נתונים, תוך שמירה על מידע רב ככל האפשר.

לכל אחד מהשיטות שמשלבות PCA השתמשנו גם ב Gaussian Mixture.

**Gaussian Mixutre** –

זהו אלגוריתם נוסף שבתוכו נשתמש גם באותן שיטות לפיצר אקסטרקשיין.

שיטה זאת בשונה מהשאר משתמש בSoft Clustering העובד בצורה דומה לHard Clustering ומחלק את כל הדאטא לCluster-ים אך כל אחד מהדאטא אינו מחולק לקבוצה אחת, אלא כל אחד מחולק באחוזים(כלומר נקודה יכולה שייכת 50 אחוז לכחול 40 אחוד לאדום ו10 לירוק).

אלגוריתם זה הוא אלגוריתם הסתברותי שמניח שכל נקודות הדאטה מיוצרות ממספר סופי של התפלגויות גאוסיאניות עם פרמטרים בלתי ידועים. מודל זה יכול להיחשב כהכללה של k-means, המשלב מידע על מבנה השונות של הנתונים ועל מרכזי הגאוסיינים הסמויים.

Gaussian mixture מממש את אלגוריתם expectation-maximization (EM) - אלגוריתם מיקסום התוחלת, היא שיטה איטרטיבית למציאת אומדני נראות מקסימלים לפרמטרים, במודלים סטטיסטיים התלויים במשתנים שלא נצפו. מודל איטרציה זה עובר בין שני שלבים: שלב התוחלת (שלב ה-E), אשר יוצר פונקציה על התוחלת של לוג הנראות המוערך באמצעות האומדן הנוכחי עבור הפרמטרים, ושלב מקסום (שלב ה-M), אשר מחשב את הפרמטרים כדי למקסם את תוחלת לוג הנראות שנמצאו בשלב התוחלת. פרמטרי הערכות אלה משמשים לאחר מכן לקבוע את ההתפלגות של המשתנים החבויים בשלב התוחלת הבא.

במילים פשוטות נתחיל את האלגוריתם ע"י בניית K(מספר הclusters) משוואות Gaussian רנדומליות בשטח.

לאחר מכן ניקח את כל הדאטא ונעשה לו Soft Clustering.

שנסיים את חלוקת הדאטא, נמחק את ה Gaussian mixture שבנינו מקודם ונחשב אותו מחדש לפי הדאטא המחולק.

כעת נחזור על החלוקה ועל בניית ה Gaussian mixtureבאיטרציות עד שנראה כי אין שינוי יותר (שינוי מינימלי) ונחזיר את התוצאות.