**מה להגיד בפרזנטציה:**

סקירה ספרותית מטרתה לבחון ולסקור מאמרים אחרים אשר דנים בנושא שלנו.

התהליך הזה מסייע לנו לקבל הבנה מקיפה של עבודות קיימות בתחום ולזהות פערים, מגבלות או אזורים לחקירה נוספת. מסגרות או מודלים קיימים יכולים להוות לנו את הבסיס למחקר.

במאמר חילקנו את הrelated word ל3 חלקים: בחלק הראשון הבאנו מאמרים שתרמו לפרויקט באמצעות טכניקות וכלים מעניינים, בחלק השני הבאנו מאמרים שונים המדברים על הבעיה שלנו של זיהוי המגדר של הדובר ובחלק השלישי הבאנו מאמרים שתרמו לפרויקט שלנו בכך שם הם אגנוסטיים לשפה.

\*\*\*\*\*\*\*\*\*עמית

**בחלק השני כמו שאמרתי הבאנו מאמרים שעסקו בנושא.**

מאמר מסוים אבחן את המאפיינים האקוסטיים של הדיבור והקולות של **ילדים** בגילאים שונים. בניסוי הראשון התקבלו ההקלטות של הילדים ובניסוי השני מבוגרים הקשיבו להקלטות הילדים ניסו לזהות את מגדר הדובר. תוצאות הניסויים הראו שמאפיינים אקוסטיים אלה הופכים ברורים יותר ככל שילדים מתבגרים. כלומר ככל שגיל הדובר גדול יותר נוצרים יותר תדרים התורמים להבחנה בין המינים. זה מרמז ששינויים בגודל הפיזי יכולים להשפיע על מאפייני הקול.

לכן החיסרון של המאמר הזה הוא שמערך הנתונים שלו מורכב רק מהקלטות של ילדים כאשר כפי שציינתי אצל ילדים יש פחות הבדל בקול בין בנים לבנות ולעומת זאת אצל מבוגרים ההבדל גדול יותר וניתן לזהות בקלות יותר את המגדר. אנחנו השתמשנו במערך הנתונים של common voice, כפי שנסביר בהמשך אשר בו ההקלטות הם של מבוגרים בלבד.

במאמר נוסף מוצג אנסמבל מוערם עבור מודל זיהוי קול מגדרי. אנסמבל מוערם היא טכניקת למידת מכונה המשלבת את התחזיות של מספר מודלים בודדים (מודלים בסיסיים) כדי לשפר את ביצועי הניבוי הכוללים.

המאמר מציג שימוש בארבעה מסווגים שונים, ביניהם SVM, .KNN,LR כלומר 4 אלגוריתמי למידה מכונה המאומנים באופן עצמאי על אותו מערך נתונים. כל מודל לומד לבצע תחזיות על סמך תכונות הקלט ומייצר סט חיזויים משלו.

הרעיון מאחורי אנסמבל מוערם הוא שעל ידי שילוב התחזיות של מודלים בסיסיים מגוונים, הוא יכול ללכוד היבטים שונים של הנתונים ולמנף את החוזקות של כל מודל בנפרד.

החיסרון של מאמר זה אנסמבל מוערם יכול גם להיות יקר יותר מבחינה חישובית ודורש בחירת דגמים קפדניים וכוונון כדי למנוע התאמת יתר.

חסרון נוסף שהוא השתמש באלגוריתמי למידת מכונה אבל CNN לא ביניהם. במאמר שלנו אנו השתמשנו בCNN בסיווג ועבור זיהוי מדר קולי שיטה זו עדיפה על SVM כיוון שCNN יכול להתמודד בצורה יעילה יותר עם מערך גדול של נתונים והוא יכול ללמוד תכונות מגוונות יותר וספציפיות לשפה הקשורות לזיהוי מגדר.

**בחלק השלישי כמו שאמרתי הבאנו מאמרים שעסקו גם בנושא האגנוסטיות.**

\*\*\*\*\*\*\*\*\*מוריה

מחקר נוסף שראינו הקשור לאגנוסטיות של השפה התמקד בקביעת המגדר של הדוברים על סמך המאפיינים האקוסטיים של הנאומים שלהם. החוקרים אספו דגימות דיבור משבע שפות שונות וניתחו את המאפיינים האקוסטיים כדי לזהות את המגדר של הדוברים. המחקר השתמש בטכניקות כריית נתונים ולמידה עמוקה, כולל אלגוריתמים של למידת מכונה כמו רגרסיה לוגיסטית, SVM, KNN ו-AdaBoost.

בדומה למחקר שלנו, הם הציגו במחקר את תוצאות ניתוח הסיווג שבוצע תוך התייחסות לאילו מאפיינים אקוסטיים של הקול, **ללא** קשר לסוג השפה, בקביעת מין הקול. הם גם ציינו שלשפות המדוברות באזורים שונים יש הגיות ואינטונציות שונות מאוד.

גם הם השתמשו במערך הנתונים של common voice אבל בשונה ממנו לקחו שבע שפות.

הם ציינו שככל שמערך הנתונים גדול ומגוון יותר, כך ההצלחה האמיתית של השיטות שהוצגו במחקרים טובה יותר.

השינוי שקיים אצלנו שהוא גם יתרון שאנחנו עושים מגוון של ניסויים בהם אנחנו כל פעם משנים את השפות שבאימון והשפות שמבחן, עשינו ניסויים שונים עם טכניקות מתקדמות יותר וכל פעם מספר שונה של שפות.

סה"כ קיימים עוד המון מאמרים בנושא הזה של זיהוי מגדר קולי כיוון שזוהי משימה חיונית המתאימה ונחוצה להרבה דברים בחיינו.

קיימים מגוון דרכים ואלגוריתמים לפתור את הבעיה הזו ולסווג אותה נכון, אבל אנו בחרנו בכמה שיטות וטכנולוגיות טובות וחזקות כדי למצוא פתרון נוסף וטוב יותר לבעיה.

ליתר בטחון:

כריית נתונים היא תהליך של גילוי דפוסים, קשרים ותובנות מכמויות גדולות של נתונים. זה כרוך בחילוץ מידע בעל ערך ולא ידוע בעבר ממערכי נתונים, לעתים קרובות במטרה לקבל החלטות מושכלות, לחזות תוצאות עתידיות או לחשוף דפוסים נסתרים.

כריית נתונים מקיפה מגוון של טכניקות ומתודולוגיות מתחומים שונים, כולל סטטיסטיקה, למידת מכונה ומערכות מסדי נתונים.

מאפיינים אקוסטיים מתייחסים לתכונות או לתכונות של צליל שניתן למדוד ולנתח. בהקשר של דיבור, מאפיינים אקוסטיים הם התכונות או הפרמטרים המתארים את הצליל המופק על ידי רמקול. מאפיינים אלה יכולים לכלול היבטים שונים של גל הקול, כגון גובה הצליל, עוצמה, משך, תוכן ספקטרלי ותבניות זמניות.

( בנוסף זיהוי מגדר באמצעות קול מצריך לעתים קרובות לימוד דפוסים מורכבים שאולי לא ניתנים להפרדה בקלות באמצעות שיטות ליניאריות כמו SVM. CNNs, עם היכולת שלהם ללמוד קשרים לא ליניאריים בין תכונות קלט, יכולים ללכוד ולדגמן טוב יותר את הדפוסים והווריאציות המורכבות של אותות דיבור הקשורים למגדר.)

SVM (Support Vector Machines): SVM הוא אלגוריתם למידה מפוקח שמוצא היפר-מישור אופטימלי להפרדת נתונים למחלקות שונות על ידי מיקסום המרווח בין מחלקות, תוך שימוש בוקטורי תמיכה כנקודות נתונים מפתח.

CNN (Convolutional Neural Networks): CNN הוא סוג של מודל למידה עמוקה המיועד לעיבוד נתונים דמויי רשת, כגון תמונות, תוך שימוש בשכבות קונבולוציוניות כדי לחלץ תכונות מקומיות וייצוגים היררכיים.

KNN (ק-השכנים הקרובים): KNN הוא אלגוריתם למידת מכונה לא פרמטרית המסווגת מופעים חדשים על סמך הצבעת הרוב של K השכנים הקרובים ביותר שלהם במרחב התכונות.

**בחלק הראשון כתבנו שני מאמרים אחד מדבר על wav2wec והשני מדבר על Mel-Spectrogram.**

* wav2wec הזה מערכת לזיהוי דוברים אשר מקבלת את קטעי הקול והופכת אותם למטריצות, מערכת זו עובדת ע"י פיקוח עצמי, זאת אומרת ללא תווית.

החידוש העיקרי של Wav2vec הוא היכולת שלו למנף כמויות גדולות של קטעי קול ללא תווית לצורך אימון מקדים, ובכך להפחית את ההסתמכות על נתונים המתומללים באופן ידני. זה מאפשר לאמן מודלים של זיהוי דיבור עבור שפות שבהם הנתונים המסומנים מוגבלים או אינם זמינים.

Wav2vec) הראתה תוצאות מבטיחות והשיגה ביצועים מתקדמים במדדים שונים של זיהוי דיבור. הוא שימש גם ביישומים כמו עוזרי קול, שירותי תמלול ומערכות מבוקרות קול.)

* מל-ספקטרוגרמה היא ייצוג של תוכן התדר של אות שמע. הוא נגזר מהספקטרוגרמה, שהיא ייצוג חזותי של ספקטרום התדרים של אות לאורך זמן. התוצאה היא ייצוג דו מימדי, כאשר ציר ה-x מייצג את הזמן וציר ה-y מייצג את התדר.

אלו שתי השיטות שבאמת השתמשנו בהם במחקר שלנו, כמו שנאמר קודם לכן לכל אחד מהם יש יתרון וחיסרון. ה wav2wecמהיר יותר בעוד לשיטת הספקטרוגרמה יש זמן ריצה ארוך יותר, אך היא נוטה להיות מדויקת יותר.