

פרוייקט

כריית נתונים

שמות מגישים:

שיר יצחק-פור 313583205

אלכס צימלמן 314347311



הקדמה:

לאחר חיפושים רבים אודות dataset שמעניין אותנו לחקור מצאנו dataset בנושא נידונים למוות בטקסט, ארה"ב. הפרוייקט יעסוק במידע אודות הנאשמים, קורבנותיהם והצהרותיהם האחרונות. תחילה, נרצה לבדוק האם יש מאפיינים ייחודיים לנאשם שנידון למוות, הרי הוא ביצע פשע נתעב ואכזרי יותר מפושעים אחרים, וקורבנותיו. נבדוק האם קיים קשר בין גילאיהם, מקום מגוריהם, מוצאם התרבותי, מינם, השכלתם וכד'. בנוסף, נרצה לבדוק ולנתח את ההצהרות האחרונות של אותם נידונים למוות.

רקע:

עונש מוות הוא אחד הנושאים השנויים במחלוקת בתחום זכויות האדם בארצות הברית. עונש המוות אושר על ידי 32 מדינות בארה"ב. טקסט מובילה את האומה במספר ההוצאות להורג מאז שעונש המוות הוחזר בשנת 1976. נאשמים יכולים למצוא עצמם מורשעים בעונש מוות משלל סיבות לדוגמא:

- רצח של אדם מתחת לגיל 10.
- רצח של יותר מאדם אחד באותה עבירה פלילית.
- רצח בעת בריחה או ניסיון בריחה ממוסד עונשין.
- רצח עובדי ציבור למשל שוטר או כבאי הפועל במילוי תפקידו החוקי.
- ועוד ...

בין השנים 1982 - 2017 הוצאו להורג 545 מואשמים בטקסט.

ניסוי:

ה-dataset כולל מידע על פושעים שהוצאו להורג על ידי מחלקת המשפטים הפלילית של טקסט מ-1982 עד 8 בנובמבר 2017 באמצעות הזרקת חומר רעיל. ה-dataset מורכב מ-545 נאשמים עם 21 פרמטרים:

- סדר ההוצאה להורג.
- שם משפחה של העבריין.
- שם פרטי של העבריין.
- מספר העבריין.
- גיל העבריין.
- מוצא של העבריין.
- מחוז של הרשעה.
- גיל של העבריין כאשר קיבל.
- רמת השכלה של עבריין.
- מחוז של העבריין.
- אם העבריין ביצע כל פשע לפני.
- מספר הנאשמים.
- מין הקורבנות.
- ההצהרה האחרונה של העבריין.

על מנת לנתח את הנתונים הללו השתמשנו בשלל אלגוריתמים וטכניקות למידה ומחקר שלמדנו הן מההרצאות והן באופן עצמאי כגון: ניתוח קורלציות שהצגנו באמצעות גרפים שונים מבוססי מפות חום, מטריצות פיזור, טבלאות וכד'. בנוסף השתמשנו באלגוריתמים במטרה לסווג לקבוצות, מדידת נכונות הממצאים שקיבלנו וניתוח טקסט חופשי.

תוצאות:

- גילינו שרוב ההצהרות הכילו מילים שקשורות לחיי משפחה ושמות של קרובים.
- גילינו שחלוקת המין של הקורבנות מתחלקת באופן יחסית שווה.
- רוב הנאשמים בעלי גוון עור לבן, דבר שאינו מפתיע מכיוון שרוב אוכלוסיית טקסס מורכבת מכאלה. להפתעתנו, חלק מתוצאות המחקר לא תאמו את השערותינו.
- ציפינו לגלות קשר ישיר בין רמת השכלתם של הנאשמים והופתענו לגלות שרובם המכריע הם בוגרי תיכון לאחר שסיימו 12 שנות לימוד ואף חלק קטן המשיך ללימודים גבוהים.
- ציפינו שלנאשמים יהיה עבר פשיעה עשיר בארועים אך תוצאות המחקר מראות כי כמחצית מן הנאשמים זוהי עבירתם הפלילית הראשונה ואין להם עבר פלילי.
- ציפינו לראות יותר ארועים בהם מעורבים נאשמים וקורבנות ממוצאים שונים אך להפתעתנו מצאנו קורלציה חיובית בין מוצא הנאשם לבין מוצא קורבנו.
- הופתענו לגלות כי כ-80% מההצהרות האחרונות היו בעלות קונוטציות חיוביות דבר אשר יכול להראות על השלמה של הנאשמים עם מצבם הנתון. ציפינו שרוב ההצרות יהיו "שליליות" מטבע הדברים - אף אדם לא רוצה למצוא עצמו במעמד שכזה שבסופו מגיע קץ חייו.

לסיכום:

מכיוון שרוב ממצאי המחקר הפתיעו אותנו ונגדו את רוב השערותינו וציפיותינו הראשונות, נורא שמחנו שבחרנו בנושא זה מסיבה זו. בעזרת כל הנתונים והממצאים אשר הגענו אליהם יצא לנו להכיר יותר לעומק את מערכת המשפט בטקסס הלא סבלנית כלפיי פשעים פליליים בין אם אלה פשעים חוזרים של אותו אזרח או פשעים ראשונים. יצא לנו לראשונה להיחשף לתחום חדש שאינו מוכר לנו ולא נתקלנו בו מעולם - דבר שהפך את הפרוייקט לעוד יותר מעניין ומשמעותי עבורנו. בנוסף, אנו מרגישים כי מעבר להרצאות והמטלות שלמדנו מהם המון, הפרוייקט נתן לכל אלו משמעות נוספת וכך תרגלנו את כל מה שלמדנו בכל הקורס ובנוסף במהלך בניית הפרוייקט למדנו באופן עצמאי עוד מספר נושאים מתחום כריית נתונים ולמידת מכונה.