**פרוייקט גמר**

StarTrace

**מחברים**: נטע לי להט, גל בן יאיר, יונתן אברהמי ועמית וייל

**מוגש לבית הספר למדעי המחשב**

**המסלול האדמאי, המכללה למנהל**

**כחלק מהדרישות האקדמאיות**

**לתואר B.Sc. במדעי המחשב**

אב התשע"ה ראשל"צ יולי 2015

**פרוייקט גמר**

StarTrace

**מחברים**: נטע לי להט, גל בן יאיר, יונתן אברהמי ועמית וייל

**המנחה המאשר**: ד"ר משה בוטמן

**בית הספר למדעי המחשב**

**המסלול האדמאי, המכללה למנהל**

אב התשע"ה ראשל"צ יולי 2015

**תוכן העניינים – לערוך בסוף**

[תקציר - 5 -](#_Toc392447285)

[הקדמה - 6 -](#_Toc392447286)

[הצדקה אקדמאית - 6 -](#_Toc392447287)

[מטרת הפרויקט - 6 -](#_Toc392447288)

[גוף הפרוייקט - 7 -](#_Toc392447289)

[1. ארכיטקטורה - 7 -](#_Toc392447290)

[2. מבנה בסיס הנתונים - 9 -](#_Toc392447291)

[3. עבודה מול ה Team Foundation Server של מיקרוסופט - 10 -](#_Toc392447292)

[אפיון המסכים - 13 -](#_Toc392447293)

[1. מסכי המערכת - 13 -](#_Toc392447294)

[מסך ניהול צוותים וחברי צוות - 13 -](#_Toc392447296)

[מסך ניהול תכונות ותתי תכונות - 14 -](#_Toc392447298)

[מסך ספרינטים - 15 -](#_Toc392447301)

[מסך צפייה בלו"ז צוותי - 16 -](#_Toc392447307)

[2. שימוש במסכי המערכת - 17 -](#_Toc392447310)

[טכנולוגיות בשימוש - 18 -](#_Toc392447320)

[קשיים בדרך ופתרונם - 20 -](#_Toc392447321)

[אפשרויות יישום לצורך הרחבת המערכת - 20 -](#_Toc392447322)

[האלגוריתם - 21 -](#_Toc392447323)

[מסקנות, יתרונות וחסרונות - 33 -](#_Toc392447324)

[תרשימים - 34 -](#_Toc392447325)

[מקורות - 35 -](#_Toc392447330)

[שימוש בספריות Open Source - 35 -](#_Toc392447331)

תקציר

בעידן הרשתות החברתיות וכמויות המידע העצומות הקיימות ברשת ניתן לצבור מידע רב היכול לשמש להצלחת שיווק ופרסום בעולם כולו.

לפני תחילת מסע פרסום ניתן להשתמש במידע זה ככלי על מנת לדעת מה דעת הרוב על מפורסם בעיר מסוימת ובכך למנוע כישלון ובזבוז כספים על פרסום לא מוצלח.

StarTrace תעזור לבחור במפורסם המתאים ביותר למסע פרסום במיקום מסוים (עיר או מדינה) בעולם בהתבסס על ניתוח סנטימנטלי של תגובות משתמשים על אותו המפורסם ברשתות חברתיות.

בממשק למשתמש יש להזין שם של מפורסם ועיר ולאחר חישוב יוצגו למשתמש אחוזי הדעות החיוביות, השליליות והניטרליות על אותו המפורסם בעיר שנבחרה.

ממשק ה- WEB ימומש ע"י טכונלוגית AngularJS ביחד עם חבילת עיצוב Bootstrap, ע"מ ליצור מראה פשוט ונקי. שרת האפליקציה ישלח בקשות אל שרת מכונת הלמידה, אשר תנתח אותן לפי אלגוריתמי הניתוח הסנטימנטלי, ויקבל חזרה את תוצאות הניתוח. מכונת הלמידה תאסוף את המידע באמצעות הממשקים השונים שמציעות הרשתות החברתיות השונות (טוויטר, פייסבוק, אינסטגרם וכו'). הקוד של מכונת הלמידה ימומש בPython 2.7

הקדמה

כיום ברשת האינטרנט קיים מידע רב אודות מפורסמים ומעריציהם. המידע אודות המפורסמים מגיע בכמויות גדולות וממקורות שונים ומגוונים. לאיסוף ועיבוד המידע פוטנציאל עסקי רב.

הגדרת עולם הבעיה - לפחות עמוד

כיום כל החברות בעולם אשר מפרסמות מוצרים לציבור הרחב צריכה מפורסם שיעמוד בראשה וייצג אותה כלפי הציבור עבור שלטי חוצות\קטלוגים\פרסומות וכדומה.

בלי פרסום זה וקמפיינים לשיווק המוצר, החברה והמוצר כנראה לא יממשו את פוטנציאל השיווק שיש להם.

מעבר לכך שיש צורך בשיווק נכון, הבחירה בפרזנטור שיהיה פני החברה עבור שיווק ופרסומים אלו היא לא פחות חשובה.

ישנם כמה סוגים של פרזנטורים:

1. פרזנטור מומחה – מקצועי בתחום הנמכר ונותן תחושה מקצועית וטובה לקונים מכיוון שמבין בתחום.
2. פרזנטור צרכן – ממליץ על המוצר בעקבות התנסות
3. ידוען\סלבריטי – שיכול לייצר אצל חלק ממעריציו גם תחושה טובה לגבי המוצר.

כאשר נרצה לבחור מפורסם להיות הפרזנטור של חברה – איך נדע במי לבחור?

כל בן אדם, כל תרבות וכל סביבה תעריץ סוגים שונים של מפורסמים כך שיכול להיווצר מצב שבסביבה מסוימת מפורסם אהוב נורא ובסביבה אחרת שנוא.

סוגיה זו יוצרת בעיה עבור חברות הפרסום שצריכות לבחור פרזנטור שיגרום לכמה שיותר אנשים ללכת אחריו ולקנות את המוצר.

הגדרת הבעיה ומה נעשה בתחום - 2 עמודים לפחות

העתקתי את זה מכתבה – צריך לשכתב ולראות מה מתאים לנו. http://www.allmarketing.co.il/?CategoryID=58&ArticleID=7500

עד כה בחינה כמותית ומעמיקה של אפקטיביות פרסומות והשוואת ההשפעה של פרזנטור כזה או אחר בקטגוריה כזו או אחרת הייתה משימה כמעט בלתי אפשרית. ניתן היה לערוך בחינה שכזו בשתי שיטות שלכל אחת מהם חסרונות משלו. הראשונה, הצגת פרסומות בפני קבוצות מיקוד, מתודולוגיה אשר לא נתנה תשובות כמותיות מובהקות לסוגיות והשנייה באמצעות הצגת פרסומות בפני עוברים ושווים במרכזי קניות, מתודולוגיה שלא אפשרה להגיע למדגם כלל ארצי מייצג.

כיום, באמצעות הפאנל האינטרנטי של חברת פאנלס, ניתן לראשונה לבחון סוגיות הנוגעות לפרסומות על פני מדגם ארצי מייצג.

בסקר שערך מכון המחקר פאנלס, עבור אגוד חברות הפרסום ובית הספר של חברות הפרסום, הוצגו בפני כ- 451 נשאלים מגוון פרסומות מ- 5 קטגוריות כאשר כל פרסומת כוללת סוג שונה של פרזנטור: ."סלב", מומחה או צרכן (סה"כ 15 פרסומות).

על כל פרסומת נשאלו מספר שאלות בנוגע ל: האהדה לפרסומת, תפיסת האמינות שלה, העניין והסקרנות שהיא יצרה לגבי המוצר/השירות המפורסם, התרומה של הפרסומת להעדפת המוצר/השירות המפורסם.

מן הסקר ניתן לראות כאשר מחפשים ליצור **אהדה** למוצר מסוים, פרזנטור שהוא "סלב" מוביל בכל הקטגוריות. כלומר, על מנת להפוך את המוצר לפופולארי, יש להשיג פרזנטור שהוא "סלב".

אך כיצד ניתן לדעת מי הוא הסלב המתאים ביותר לשיווק המוצר אותו אנו רוצים להפוך לפופולארי?

נראה שאין דרך פשוטה למצוא מי מתאים מכיוון שאין את הכלים המתאימים למצוא איזה "סלב" הוא האהוד ביותר בסביבה מסוימת. ניתן להסתמך על דברים כמו רמת חשיפה או הופעות במדורי הרכילות, אך לעיתים דברים אלה עלולים להתפספס ובנוסף חסר מיקוד עבור ה"סלב" ברמת המיקרו, איפה הוא יותר פופולארי? בקרב אילו גילאים? האם יותר פופולארי בקרב נשים או גברים?

פתרון נוסף שקיים בעולם הדיגיטלי הוא Google Trends ו-Facebook Insights.

Google Trends יכול לתת מידע על כמה מחפשים ברחבי הרשת נושא מסוים או אדם מסוים ובאילו מקומות. Facebook Insights מציע לבעלי עמוד מסוים לקבל דוחות על כמות התנועה בעמוד שלהם ובכך יכול לספר לבעלי עמוד של "סלב" מסוים מהי רמת הפופולאריות שלו.

פתרון - הסבר על הפתרון המוצע וכיצד שונה מאחרים

מתוך שלושת הסוגים של פרזנטורים שדיברנו עליהם (צרכן\מומחה\ידוען), StarTrace באה לתת פתרון לסוג השלישי- בחינת אהדת הידוענים.

ניתן גם לתת פתרון עבור שני הסוגים הראשונים במידה והם גם דמויות הידועות לציבור.

מערכת StarTrace בונה על חוכמת ההמונים, ומחפשת שם את התשובה לשאלה שנשאלה בסוף הסעיף הקודם, מהי רמת הפופולאריות של "סלב" מסוים? היא עושה זאת באמצעות פניה אל הרשתות החברתיות בכדי להבין את רמת האהדה שלהם בקרב הציבור. באמצעות ניתוח סטאטוסים, Tweetים, תגובות ודיווחים על אותו ה"סלב", נוכל ללמוד מהי רמת האהדה שלו ברחבי הרשת. את רמת האהדה נוכל לאחר מכן להציע לחברות פרסום שונות אשר מחפשות את הפרזנטור הבא עבור המוצר שלהם. את החיפוש שלנו ניתן לבצע לפי מיקום ורשת חברתית ובהמשך תינתן האופציה לחפש גם לפי גילאים ומין.

הפתרון שלנו שונה מזה של Google Trends בכך שהוא יכול לסווג את ה"רעש" שהאדם אותו מחפשים עושה ברשת, בעוד ש-Google Trends בסך הכל יגיד **כמה** מחפשים אותו. מדדי הסיווג שלנו הם "חיובי", "שלילי" ו-"נייטרלי". בנוסף, אנו שונים מ-Facebook Insights בכך שאנחנו מאפשרים גם לאנשים שהם אינם מנהלי העמוד של ה"סלב"(ובכך למעשה מקורבים ברמה כלשהי אליו) לדעת מהי חוות הדעת עליו ברחבי העולם.

הצדקה אקדמאית

כדי להסיק תובנות בעלות פוטנציאל עסקי מתוך מאגרי הנתונים הנאספים מרשתות חברתיות על מפורסמים נשתמש באלגוריתמי למידה אשר נלמדו בקורס למידה חישובית וכריית נתונים. אנו עבדנו עם שני אלגוריתמים מרכזיים:

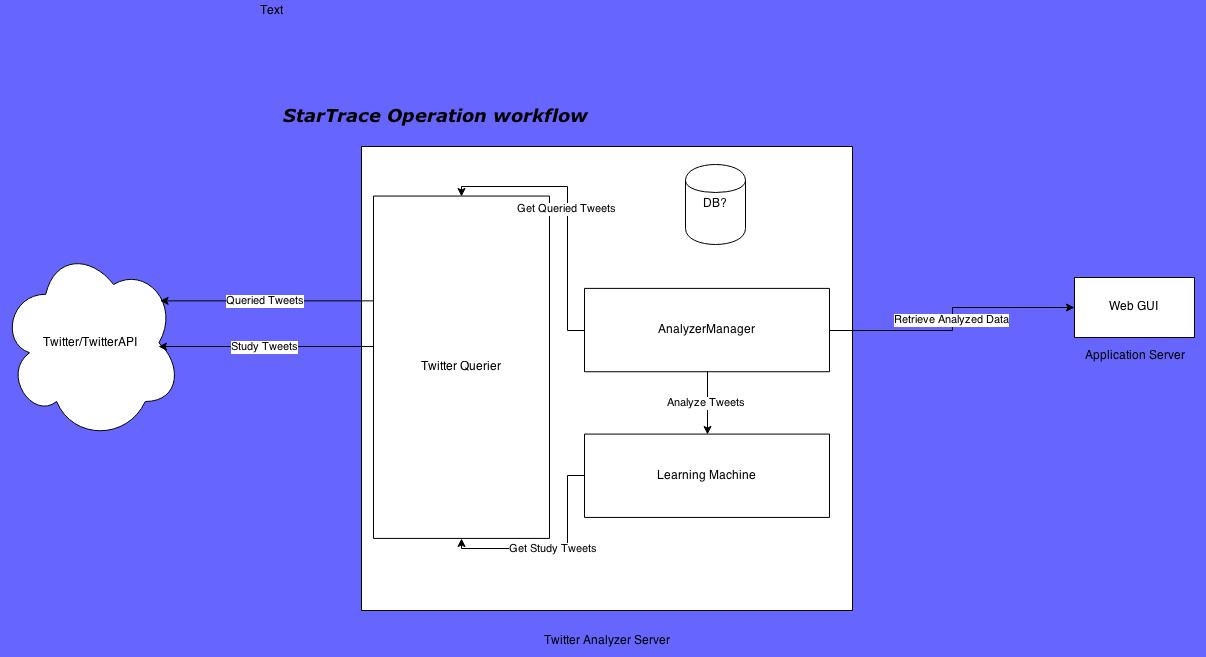
וכאן עמית כותבת על Naive Bayes ו-SVM לפחות

מטרת הפרויקט

מטרת הפרויקט היא מציאת תובנות עסקיות בעזרת אלגוריתמי למידה על בסיס מידע רב שנאסף מרחבי האינטרנט. כדי לקבל החלטות עסקיות בתחום הפרסום נדרש להשתמש במידע של דעת הקהל הרחב על מפורסמים כדי שתהליך הפרסום יהיה מוצלח. נבנה מערכת שמספקת ידע לגבי דעת הקהל על מפורסמים על פי ערים בעולם על בסיס מידע..

תרשים בלוקים\מודולים כללי של המערכת

יוני – זכור לי שעשית פעם תרשים של הארכיטקטורה, נראה לי שצריך לשפץ אותו ולהוסיף לו עוד תרשימים



DR

SRS

מתודולוגיות וכלי פיתוח, חומרה ותוכנה

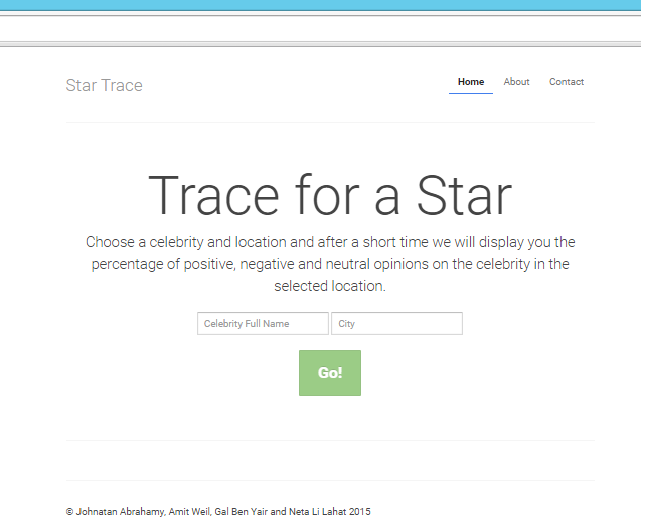
המערכת מחולקת לשני חלקים – צד שרת וצד לקוח. צד השרת פותח לחלוטין ב-Python על מנת להשיג זריזות וגמישות בעבודה עם הממשקים. על מנת להפוך מיקום משם של מקום לקואורדינציות השתמשנו בשירות רשת של הכנס שם כאן. את ממשק ה-REST פיתחנו באמצעות טכנולוגייה מבוססת Python בשם "הכנס שם כאן". נטע לי תרחיבי על צד לקוח

המערכת - כולל תמונות מסך של קלט ופלט המערכת

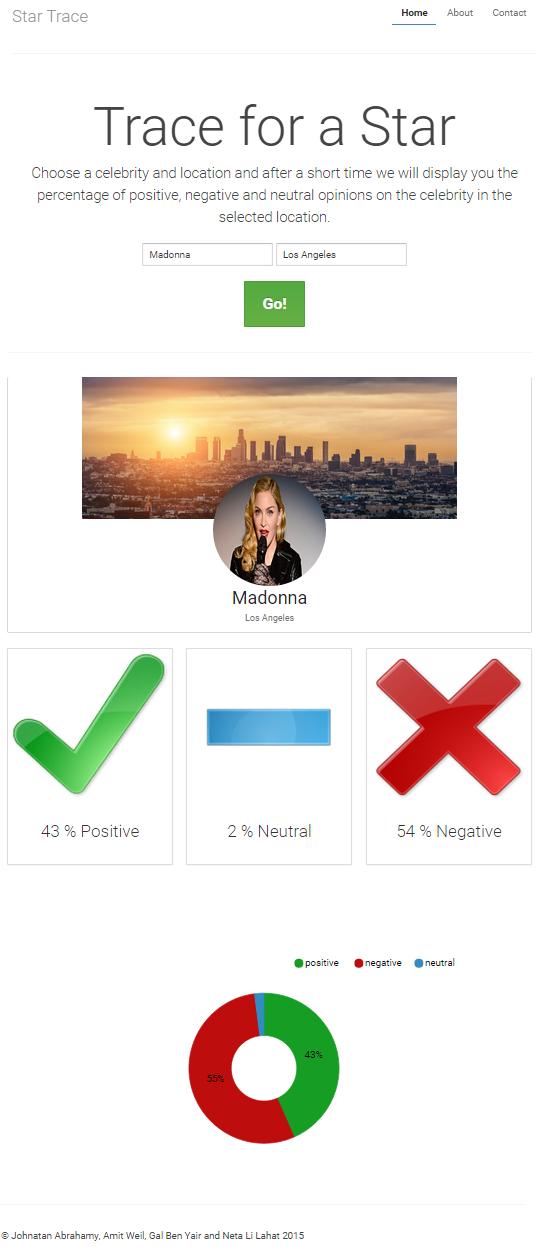
מסך הבית של המערכת:

התחברות המשתמש לאתר תביא אותו ישר לצפיהה בדף הHome.

* מאפשר הזנה של שם מפורסם ועיר בעולם עליהם ירצה המשתמש להריץ את סקירת התגובות.



לאחר הזנת הנתונים ולחיצה על "GO" מופעל האלגוריתם והמשתמש מקבל את תוצאות המערכת.

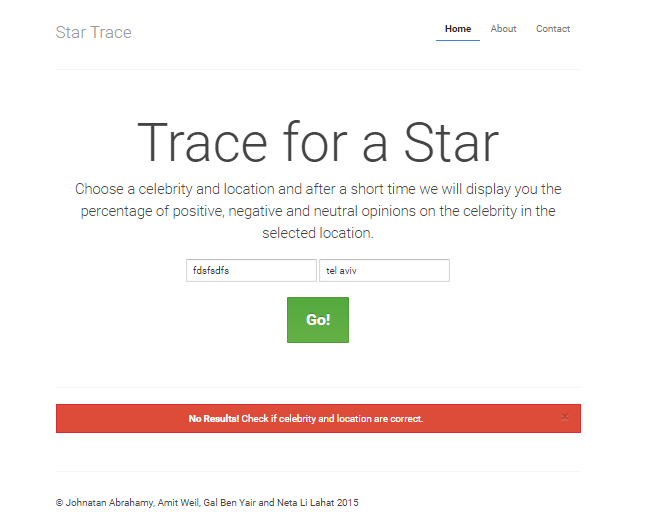


* מוצגת תמונה המפורסם ותמונה המקושרת לעיר שנבחרה
* התוצאות מתחלקות ל3 קטגוריות:

1. חיובי
2. שלילי
3. ניטרלי

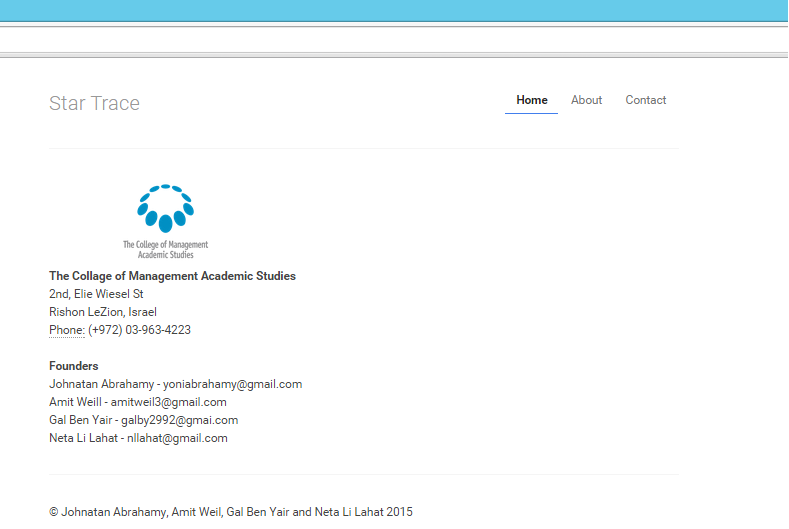
* ניתן לראות את התוצאות גם בצורת גרף.

במידה והמשתמש הזין נתונים (עיר\מפורסם) לא קיימים, המערכת לא תחזיר נתונים ותתקבל ההודעה הבאה:



מסך יצירת הקשר

בעת לחיצה על כפתור הcontact בתפריט המשתמש יגיע לדף יצירת הקשר הבא :



מסך הabout

מכיל פרטים כלליים על המוצר StarTrace .

מסך זה מציג גם את היכולת העסקית של המוצר עבור חברות שמחפשות מפורסמים.

משימות על פי זמנים והסבר על עמידה בזמנים

* בדה"ת לטכנולוגיות ודרכי איסוף נתונים מרשתות חברתיות שונות- ינואר 15
* בחירות טכנולוגיות ותחילת מימוש – פברואר 15
* פיתוח אופן הלמידה ובחינת פלט האלגוריתם – מרץ 15
* פיתוח ממשק למשתמש – אפריל 15
* הכנת תקצירים, ברושור ופוסטר לפרויקט - מאי 15
* הצגת הפרויקט – יוני 15
* ספר פרויקט – יולי 15

צורת העבודה שבה עבד הצוות למימוש הפרויקט הייתה חלוקת משימות בין חברי הצוות כדי שנעמוד בלוח הזמנים הנדרש לפרויקט, עמדנו בזמני העבודה שנדרשו.

קשיים במהלך הפרויקט וכיצד נפתרו

הקושי המרכזי במהלך הפרויקט היה למצוא API ידידותי אשר יאפשר לנו להוציא מידע על אנשים מרחבי הרשתות החברתיות. ראשית ניסינו לעבוד מול Facebook, אבל לאחר שגילינו שהם ביצעו שינויים ב-API שלהם שמונעים הוצאת מידע בסיסי על משתמשים החלטנו לעשות שינוי כיוון (Pivotting) ולהוציא מידע מ-Twitter במקום. כעת נתקלנו ב-API ידידותי יותר המאפשר לאחזר Tweetים ממשתמשים שונים ובנוסף לקבל Tweetים לפי מיקום. לאחר שאחזרנו את ה-Tweetים כל מה שנותר הוא לנתח אותם ולייחצן את התוצאת באמצעות ממשק REST מעל HTTP. קושי נוסף שנתקלנו בו הוא שה-API של Twiiter מאפשר להוציא רק מספר מסוים של Tweetים מתקופה שלא תעלה על חודש. פתרון שחשבנו על מנת לפתור את הבעיה הזו היא לשמור את הTweetים במסד נתונים משלנו, ובהמשך ליצור ממשק המבצע שילוב בין אחזור מידע מ-Twitter לבין עבודה מול מסד הנתונים שלנו על מנת להפיק ביצועי למידה טובים יותר ולהגיע לתוצאות איכותיות יותר.

ביבליוגרפיה