|  |  |  |  |
| --- | --- | --- | --- |
| שם מלא | תעודת זהות | מייל | שם המתרגל |
| סהר אגרנוב | 316263045 | [Agranov96@gmail.com](mailto:Agranov96@gmail.com) | אופיר הררה |
| אביחי גל אור | 207051848 | [Avichai504@gmail.com](mailto:Avichai504@gmail.com) | שי זיו |

**תאריך הגשת הצעת פרויקט: 15.5.23**

**שם הפרויקט: Prediction of High Tech Salary in the Global Market**

**שאלת המחקר:** האם ניתן לחזות שכר שנתי בהינתן כישורי המתכנת?

- ניסוח שאלת המחקר:

\* ניסוח תקין: שאלת המחקר צריכה להתמקד בחיזוי של משתנה אחד (עמודה) בדאטה שלכם.

\* ניסוח לא תקין: נא לא לנסח שאלת מחקר בצורה הבאה: "האם קיים קשר בין משתנים (עמודות)".

האם ניתן לחזות לפי גיל האם

**הסבר כללי על הפרויקט (עד 4 משפטים):**

**הפרויקט שלנו מתרכז במאפיינים השונים של קריירה בשוק ההיי-טק העולמי.  
אנו ננתח נתונים של עשרות אלפי משרות שונות, במטרה לחזות את השכר השנתי של עובד בהינתן כישוריו, ניסיונו, סוג ההשכלה שלו, ועוד.  
לבסוף, התוכנית שלנו תשלב מספר מצומצם של משתנים מסבירים איכותיים המשפיעים על השכר השנתי של עובד, ובכך תוכל לתת מענה למתכנתים בתחילת דרכם לבנות לעצמם קריירה מותאמת לציפיות השכר שלהם, ולשאר המאפיינים הרלוונטיים.**

**מקורות מידע + קישורים:** מקור המידע שלנו הוא האתר **Glassdoor** . האתר מכיל מאות אלפי משרות עבור כל תחום בשוק ההייטק. כל משרה מכילה דרישות לתפקיד, שכר שנתי, ועוד המון נתונים שעל בסיסם אנו מנתחים את שוק העבודה בתעשייה.

**https://www.glassdoor.com/Job/data-scientist-jobs-SRCH\_KO0,14.htm?context=Jobs&clickSource=searchBox**

**שיטת הרכשה: Selenium**

- שימו לב:

\* Crawling = 15 Points

\* API (Only) = 5 Points

\* Crawling + API = 15 Points

**מה מייצגת כל שורה בדאטה סט?**

כל שורה מייצגת משרה אחת. הנתונים על המשרה הורכשו ופוענחו מפוסטים של משרות לחיפוש עבודה העולמי בהייטק.

חשוב לאמר, עבור כל משרה אנו מנתחים את המידע באופן קפדני. לאחר הרכשת הנתונים הגולמיים, התהליך ממשיך למספר פונקציות שעובדות באופן בלתי תלוי, ולכל פונקציה מטרה אחת – לפענח מהו המידע הקונקרטי שבדרישות המשרה (שנות ניסיון, ידע בתוכנות מסוימות, ולא נתונים כמו "בעל מוטיבציה רבה...")

**לצורך נוחות הקורא, מצורפת טבלה המתארת את תפקידה של כל אחת מהעמודות. גילוי נאות – בשלב זה אין אנו יודעים אילו מהנתונים יהיו מסבירים לשכר השנתי ואיזה פחות, אנו מתרכזים בהרכשת כל הנתונים האפשריים. בהמשך הפרויקט, לאחר** EDA **אנו ננפה את המשתנים שאינם משפיעים על השכר השנתי ובכך נמנע בסיבוכיות זמן גבוהה מידי.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Annual Salary** | **Experience** | **Job Title** | **Education** | **Revenue Scale** | **Company Old** | **Company Size Scale** | **Is Remote** | **Rating** | **Career Opportunities** |  |  |
| **זוהי עמודת החיזוי, תווית ה-y שלנו. זהו ערך רציף המשקף את השכר השנתי בדולרים.** | **שנות הניסיון הדרושות עבור כל משרה** | **תיאור המשרה (2-6 מילים)** | **משתנה קטגוריאלי:**  **השכלה תיכונית, תואר ראשון, תואר שני, דוקטורט.** | **שווי החברה בדולרים המחולק לטווח יחסי בין 1-9.** | **וותק החברה בשוק** | **גודל החברה (מספר אנשים) המחולק באופן יחסי בין 1-9** | **משתנה בוליאני (האם העבודה מתקיימת באופן מקוון או פיזי)** | **ממוצע של כל הדירוגים שניתנו ע"י עובדי החברה** | **משתה רציף בין 0 ל 5 הניתן ע"י דירוג עובדי החברה** |  |  |
| **Comp & Benefits** | **Culture & Values** | **Senior Management** | **Work Life Balance** | **Sector** | **Type of Ownership** |  |  |  |  |  |  |
| **משתה רציף בין 0 ל 5 הניתן ע"י דירוג עובדי החברה** | **משתה רציף בין 0 ל 5 הניתן ע"י דירוג עובדי החברה** | **משתה רציף בין 0 ל 5 הניתן ע"י דירוג עובדי החברה** | **משתה רציף בין 0 ל 5 הניתן ע"י דירוג עובדי החברה** | **תחום ספציפי בתוך החברה (משתנה מסביר לשכר העובד!!!)** | **סוג בעלות החברה (פרטית\ציבורית\עמותה ללא רווח...)** |  |  |  |  |  |  |

**סוג למידת מכונה (בהתאם לשאלת המחקר) :**

מכיוון שאנו מתעסקים בחיזוי משתנה רציף (שכר שנתי), סוג למידת המכונה שאנו נשתמש הוא רגרסיה.

**במידה ובחרתם ב למידה מונחית, מהי עמודת המטרה?**

- מונחית – רגרסיה/קלסיפיקציה.

- לא מונחית

**בחרנו בלמידה מונחית, רגרסיה. עמודת המטרה היא כאמור, חיזוי השכר השנתי.**

הנחיות הגשת הצעת הפרויקט:

1 .יש להעלות את טופס הצעת הפרויקט לדרייב המצורף של המתרגל הרלוונטי עד לתאריך ה- 18.5.23.

- מומלץ להקדים ולהגיש את טופס הצעת פרויקט עוד לפני תאריך זה, כיוון שיכול להיווצר מצב שבו לא תקבלו אישור לטופס שהגשתם, ותצטרכו לתקן.

**תאריך הגשת הצעת הפרויקט הוא נכון לגבי התאריך שבו העליתם את הטופס שקיבל אישור סופי מהמתרגל.**

2 .הגשות מאוחרות של הצעת המחקר יישאו בקנס של חצי נקודה ליום ויתאפשרו עד לתאריך ה-1.6.23

**שימו לב:** ניתן לראות את תאריכי ההעלאה של טפסי הצעות הפרויקט, ולכן הימנעו ממצב לא נעים.