**מבוא למדעי הנתונים**

**פרויקט גמר**

דניאל כהן 208423715

[Danielicohen1971@gmail.com](mailto:Danielicohen1971@gmail.com)

עמית כהן 316439892

[Amitcohen2097@gmail.com](https://d.docs.live.net/fb40e300dc5a730e/שולחן%20העבודה/Amitcohen2097@gmail.com)

**zap רקע כללי:**

zap השוואת מחירים הוא אתר השוואת המחירים למוצרים ולשירותים הגדול והמוביל בישראל, בו מבקרים מדי חודש כ-10 מיליון גולשים, האתר הוביל בשנים האחרונות מהפכה צרכנית ושינוי של ממש בתרבות הקנייה באינטרנט.

**שאלת מחקר:**

האם ניתן לחזות מחירי מזון לכלבים וחתולים על פי מאפיינים ונתונים?

**הגדרת הבעיה:**

שאלת המחקר מתעסקת ביכולת לחזות את מחירי המזון לכלבים וחתולים באמצעות ניתוח ועיבוד נתונים וטכניקת למידת מכונה. השאלה מנסה לחזות את הקשר בין מאפיינים מסוימים למחירי מזון, ובכך לעזור לצרכן לשלם את המחיר הנמוך ביותר.

**נושא הפרויקט:**

הנושא נבחר ממניע אישי, אנחנו בעלות כלבים ולאינטרס האישי היינו שמחות לדעת איך לחזות מחירים של מוצרים כדי לקנות בחוכמה.  
התוצאות שניתן להפיק ממחקר זה :

**בפן הצרכני:** המחקר עשוי לספק עזרה לבעלי כלבים וחתולים בהבנת המחירים והמאפיינים המשפיעים עליהם. בעלי חיות מחמד יכולים להשתמש במידע הנצבר על מחירי המזון כדי לקבל מידע נוסף ולקבל החלטות טובות יותר לגביו.

**בפן השיווקי:** המחקר עשוי לספק יתרון משמעותי לענף המזון לחיות מחמד ולתעשיית החיות המחמד כולה. על ידי הבנת הקשרים בין המאפיינים למחירים.

**שלבי עבודה בפרויקט:**

**איסוף נתונים:** ביצועscraping&crawling לאתר zap

להלן כמה דוגמאות מרשימת המאפיינים:

1. סוג (כלבים, חתולים, גורים)
2. מותג
3. תכולה (לפי ק"ג)
4. גזע
5. סוג מזון (יבש,רטוב)
6. טעם

הנתונים הנ"ל נלקחו מהאתר zap תחת הלשונית מזון לכלבים וחתולים.

להלן הקישור:

<https://www.zap.co.il/paramselect.aspx?sog=an-dogfood>

**ניקוי ועיבוד נתונים:**

1. בדיקת נתונים וזיהוי ערכים חסרים.
2. בדיקה האם יש מספר נתונים כפולים במאגר הנתונים.
3. בדיקה וטיפול בערכים חריגים.

**חקירת נתונים:**

1. הבנת המשתנים השונים בנתונים-טווחים, ערכים חסרים וערכים חריגים.
2. ביצוע ניתוחים כגון-מדידות ממוצעות, השוואות בין קבוצות.

**מודלים של למידת מכונה:**בשביל לחזות את מחירי המזון לכלבים וחתולים נשתמש במודלים של רגרסיה-רגרסיה ליניארית.  
ייתכן שנשתמש גם במודלים שונים כמו עצי החלטה, ורשתות נוירונים לצורך השוואה ובדיקת יעילות.

**אימון מודלים:**

1. איסוף נתונים המתאימים לשאלת המחקר שלנו.
2. בניית המודל שאותו בחרנו.