

## המחלקה להנדסת תוכנה

### פרויקט גמר –

אגירה וסיווג של מסלולי גופים פיזיקליים

באמצעות למידת מכונה

## Storage and Classification of Trajectories via Machine Learning Methods

מאת

אביתר גולן, 203311733

מנחה אקדמי: דר' ראובן יגל	אישור:	תאריך:
אחראי תעשייתי: דר' שי אקו	אישור:	תאריך:
רכז הפרויקטים: דר' אסף שפנייר	אישור:	תאריך:

מערכות ניהול הפרויקט:

#	מערכת	מיקום
1	מאגר קוד	<a href="https://github.com/EvjaG/FinalProject">https://github.com/EvjaG/FinalProject</a>
2	קישור ליומן	<a href="https://mail313575.monday.com/boards/2070942680/">https://mail313575.monday.com/boards/2070942680/</a>
3	קישור לסרטון	<a href="#">לחצו כאן לפתיחה</a>

סוג הפרויקט	תעשייתי, חברת hi-tech
פרויקט מח"ר	לא
פרויקט ממשיך	לא
פרויקט זוגי:	לא

כלי פיתוח

שפות	Python, Java
DB	MongoDB
יצירת תרשימים לדוחות	StarUML

### מילון מונחים:

**למידת מכונה** – תחום במדעי המחשב העוסק בפיתוח אלגוריתמים המאפשרים למחשב לעבד נתונים מתוך דוגמאות ומכך להסיק מסקנות לגבי מידע חדש שלא נתקל בו בעבר.

**רשת נוירונים מלאכותית** – משפחת מודלים בתחום למידת מכונה שפותחה בהשראת תהליך הלמידה המתרחש במוח אנושי. מתחלקת לשלב שבו הרשת מתאמנת על ידי קבלת דוגמאות, ושלב שבו הרשת כבר 'מאומנת' ויכולה לתת ערכים לדוגמאות.

**משקולות** – בחלק מהאלגוריתמים של למידת מכונה, אנו רוצים 'לתקן' תופעות מסויימות על מנת למנוע או לקדם את השפעתן. עושים זאת ע"י הכפלת הערך במשתנה שנקרא **משקולת**, על מנת לשנות את השפעתו של הערך על החישוב בשלבי הלמידה וההערכה. המשקולות יכולות להתעדכן לפי הדוגמאות.

**מסלול** – בפרויקט זה, מסלול הוא רשימה של נקודות במרחב המתארות את תנועתו של גוף, כאשר הנקודה הראשונה מייצגת את תחילת ההקלטה  $t = 0$ , וכל נקודה  $i > 0$  נוספת מייצגת נקודת זמן  $t_{i+1}$ . פורמט ייצוג המסלול בפרויקט הוא בקבצי CSV, כאשר כל שורה מייצגת נקודות  $x, y, z$ . בעתיד אולי יוספו פרטים נוספים לקבצי המסלול (כגון מהירויות ותאוצות, גודל הגוף, מבנה וכו').

## תוכן עניינים

4	0. מבוא .....
4	1. תיאור הבעיה .....
4	2. סקירת עבודות דומות .....
4	3. תיאור הפתרון .....
4	דרישות ואפיון הפתרון .....
4	האתגרים של הפתרון מבחינת הנדסת תוכנה .....
5	המשך הפרק תיאור הפתרון .....
5	רכיבים .....
5	תהליכים ומצבי המערכת .....
6	תיאור הכלים המשמשים לפתרון .....
7	4. סטאטוס הפרויקט .....
7	5. תכנית בדיקות .....
7	בדיקות יחידה .....
7	בדיקות אינטגרציה .....
8	בדיקות מערכת .....
9	6. נספחים .....
9	א. טבלת סיכונים .....
10	ב. טבלת דרישות .....
10	ג. תכנון הפרויקט .....
10	ד. סקר שוק .....
13	ה. תרשימים .....
15	ו. רשימת ספרות \ ביבליוגרפיה .....

## 0. מבוא

כיום מוערך שיש בעולם כ-47,840 כלי טיס צבאיים (WMDA) בשימוש, יותר מ-4,500 לוויינים פעילים המרחפים במסלול מעל האטמוספירה ואין-ספור כלי נשק המסוגלים לנוע באוויר או בחלל. במדינת ישראל ישנו צורך תמידי לשמור על בטחונה, לרבות בטחונה האווירי. במשך שנים רבות נאספו ותועדו מסלולי גופים רבים שנעו שתוך ומחוץ לשטחה האווירי של מדינת ישראל. הלקוח בפרויקט הוא גוף בטחוני המעוניין במעקב אחרי גופים אלו.

## 1. תיאור הבעיה

הבעיה – המצב הקיים כיום: המעקב כיום הוא אחרי גופים ספציפיים בעזרת רדאר הוא מאד מוגבל, והחובה שעל מטוסים להשתמש במערכות בקרת טיסה (לדוג' ADS-B) אינה חלה על כלים צבאיים ועל כלים אחרים כגון רחפנים כרגע חל איסור על השימוש בטכנולוגיות החדשות על מנת לא להעמיס על תשתיות הרדיו. הצפי בעתיד הקרוב הוא הצורך במעקב אחרי גופים רבים לא ידועים בעזרת מערכת אוטומטית למעקב אחר גופים, זאת באמצעות חיישנים כלליים (כדוג' רדאר עם רגישות גבוהה יותר מראדרים הקיימים כיום).

## 2. סקירת עבודות דומות

כחלק מהפרויקט, נחקרות מערכות דומות שבהם משתמשים בחברות ליהוי וקטלוג מסלולים בצורה אוטומטית (לדוג' ברכבים אוטונומיים), בשלב זה בוצע בעיקר סקר שוק לגבי סוגי מסדי נתונים שיכולים להתאים לכמות גדולה של נתונים מסוג זה. השוואת מסדי הנתונים השונים נמצאת בנספחים. בנוסף, בהמשך המחקר נשווה בין סוגי למידה שונים, כולל למידה עמוקה ולמידת מכונה.

בנוסף, בוצע מחקר (Bosson, Nikoleris 2018) לבדיקת היתכנות האפשרות לחזות מסלולי נחיתה של מטוסים אזרחיים אשר גילה כי ניתן להגיע לרמת דיוק של 90% בסיווג עבור אימון באורך של 40 שניות עם מירב המודלים של למידה חישובית, ועבור זמני חישוב ארוכים יותר, ייתכן להגיע אף לתוצאות גבוהות יותר עם רשתות נוירונים.

מחקר אחר (Sakyi-Gyinae 2019) ניסה לשפר את מערכות הבקרה לבטיחות אווירית בעזרת שיפור המודלים לחיזוי תעבורה אווירית, ומצא ששימוש במודל LSTM משפר את יכולות החיזוי על פני מודלים קיימים.

ניתן למצוא קישור למחקרים ב**רשימת הספרות**.

## 3. תיאור הפתרון

### דרישות ואפיון הפתרון

הקמת מערכת הכוללת בסיס נתונים ומערכת לאגירת נתונים בזמן-אמת מחיישנים הקולטים גופים פיסיקליים הנעים במרחב, הכולל תיעוד מסלולם ותכונותיהם. המאגר יאחסן ויקטלג את הנתונים בעזרת הכלים של למידת מכונה ויאפשר זיהוי הגוף בהשוואה לגופים אחרים. נטמיע מס' אלגוריתמים של למידת מכונה ונעזר בהם על מנת למדוד את ביצועי המסלול של גוף חדש וסיווגו.

### האתגרים של הפתרון מבחינת הנדסת תוכנה

- בניית מערכת שתמדוד את ביצועי האלגוריתמים ותשווה ביניהם.
- אין מסלולים אמיתיים - מידע קשה להשגה (כרגע יש לנו מעט ממנו) ולכן אין אפשרות להסתמך על מידע אמיתי.
- התאמה של נתונים מחיישנים וממאגרי מידע קיימים לצורה הניתנת ללמידה בעזרת למידת מכונה.
- בניית מערכת הלמידה באופן שבו מסלולים באורכים שונים יוכלו להתאים למערכת לומדת, היות ומס' הכניסות מוגבל לגודל מסויים.
- הקמת מערכת מבוזרת שתוכל לאגור ולקטלג את המידע, ולמשוך אותו בשלב יותר מתקדם.

תיכון הפתרון מכיל מס' רכיבים, ובהם:

- מחולל מסלולים, לצורך יצירת מסלולים בסיסיים עליהם נלמד בהמשך ומערכת הכותבת מסלולים אלו לקבצי תיעוד וקבצים ללמידה.

- מערכת למידה – ככל הנראה רשת נוירונים.
- מסד נתונים – נוכל לאחסן בו את כל המסלולים והמידע שקשור ללמידת המכונה לצורך שליפה מהירה במידת הצורך (לדוג' לזיהוי אובייקט הנע במהירות במרחב).

## המשך הפרק תיאור הפתרון

נא לפנות לנספחים לטובת תרשימי זרימה המתארים את פעולות המערכת המרכזיות.

### רכיבים

המערכת בנויה משני חלקים עיקריים;

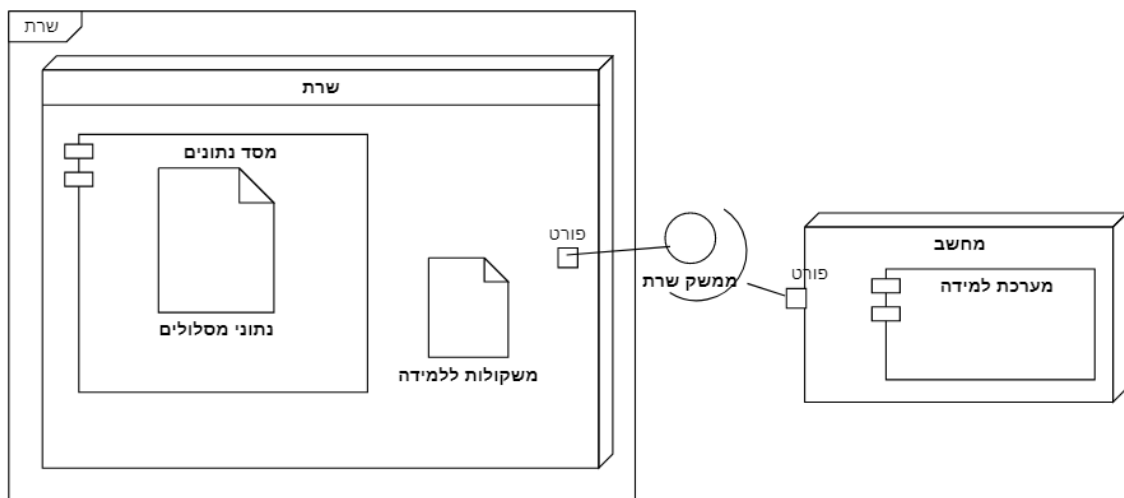
#### מכונת למידה

יכיל את תוכנת הלמידה (כקובץ הרצה) עבור מערכת הסיווג, יוכל לפנות לשרת כדי להוריד את המשקולות המעודכנים במידת הצורך, ולאחסן מסלולים חדשים עם או בלי הסיווג.

#### שרת נתונים

השרת יכיל את מסד הנתונים, המשקולות המעודכנות ללמידה. יספק ממשק משתמש (REST)

תרשים רכיבים המתאר את מבנה המערכת ממבט עליון:



### תהליכים ומצבי המערכת

המערכת תכלול שלשה מצבים עיקריים:

- סיווג:

- **סיווג של נתונים חדשים** - נתונים שייכנסו מתוך חיישנים חיצוניים יעברו סיווג בזמן אמת על ידי המכונה הלומדת
- **סיווג של מקבצי נתונים** - נניח שיוכנסו נתונים למכונה ממקור חיצוני (לדוג' מנהל הוסיף מסלולים שנרכשו מחברה חיצונית) שלא נלמדו עדיין, אנו נבצע עליהם תהליך סיווג במידה ואין ידע על איזה סוג מסלולים אלו.

- **למידה**: המערכת תיקח את כל הנתונים שיש להם סיווג חיצוני (כלומר ידוע איזה סוג מסלול זה ופרטים נוספים קיימים עליו במידת האפשר), ותבצע תהליך למידה מאפס, עם שאיפה לרמת דיוק שתיקבע ע"י מנהל המערכת. הלמידה תתבצע אוטומטית אחת לתקופה במידה ונוספו נתונים מתאימים למסד מאז הלמידה האחרונה, או ידנית ע"י מנהל מערכת.

#### תיאור הכלים המשמשים לפתרון

בפרייקט השתמשתי עד כה ב-Python לכתובת הקוד, ובספריות החיצוניות הבאות:

הספרייה	הסיבה שהשתמשתי
NumPy	חישוב ויצירת המסלולים, פונקציות מתמטיות נוחות
Pandas	בדיקה של נכונות החישובים של המסלולים, חישוב מהירויות, הצגה נוחה יותר להדפסה על המסך
OpenCV	המרת מסלולים ממערכים לתמונות, לטובת בדיקה האם רשתות קונבולוציה יתאימו ללמידה של המסלולים
Matplotlib	המחשת המסלולים בעזרת אנימציה

בנוסף, נעשו בדיקות על ביזור השרת למס' מחשבים בעזרת מערכת הקבצים Hadoop ומנוע העיבוד המקבילי Spark, של ארגון Apache, אך בשלב זה הנושא הוקפא למטרת התקדמות עם נושא הסיווג.

#### 4. סטאטוס הפרויקט

- ☒ יצירת מחולל מסלולים, כולל כתיבה לקבצי CSV ו PNG
- ☒ יצירת מודול לקריאת ה CSV ופרסור (לדו' מציאת המהירות בכיוון X בזמן T)
- ☒ בחירת מסד נתונים לעבודה ראשונית (MongoDB)
- ☒ חקר על סוגי אלגוריתמי הלמידה השונים לטובת הפרויקט
- ☐ יצירת מכונה לומדת לסיווג המסלולים
- ☐ ביצוע ניסויים ראשוניים ללמידה, ללא בדיקה
- ☐ בדיקת המסלולים במחלקת test
- ☐ כתיבת מודול לאחסון ושלפת המסלולים ממסד הנתונים

#### 5. תכנית בדיקות

על המערכת יבוצעו שלוש סוגי בדיקות, אותן אתאר בפירוט להלן:

##### בדיקות יחידה

מס'	תיאור	תנאי קדם	צעדים	דאטה	צפי	תוצאה	עבר/נכשל
1	כתיבת מסלולים	התיקיה data צריכה להיות ריקה מתוכן	1. בחירת מס' המסלולים (X) 2. בחירת Y נקודות למסלול 3. הרצה 4. בדיקה שיש X מסלולים, והחלוקה בין תיקיות train ו- test בערך 80-20	אין	קבלת $X \cdot 0.8$ קבצים בתיקיה $X \cdot 0.2$ , train קבצים בתיקיה test, כל קובץ עם שורות Y		
2	אחסון נתונים במסד	המסד צריך להיות ריק טרם תחילת הבדיקה	1. בחירה של איזשהו סט של דאטה אקראי 2. הכנסת הנתונים למסד 3. משיכת הנתונים בחזרה מהמסד 4. השוואה של הנתונים מול המקור לשגיאות	יווצר על-ידי המערכת, גודל וכמות יקבעו מראש ע"י הבודק	לא יהיו שגיאות		

בהמשך הדרך ייתכן כי נחליט שנכון כי מערכת קבצים מבוזרת (כדוג' Hadoop) תתאים למטרת הפרויקט. אם כן, נצטרך לבנות מערכת להזנת המסלולים לתוך המסד שישב על גבי מערכת הקבצים המבוזרת בצורה נוחה ומהירה, ולבדוק גם אותו.

##### בדיקות אינטגרציה

- שלפה של מסלול חדש ממסד הנתונים ולמידה עליו, ללא מידע חדש מהחיישנים.

-

מס'	תיאור	תנאי קדם	צעדים	דאטה	צפי	תוצאה	עבר/נכשל
1	שלפת מסלולים מהמסד	המערכת צריכה להיות	1. לכתוב X מסלולים עם תוויות ידועות	אין	קבלת דיוק ברמה A לפחות $(0 < A \leq 1)$		

		ייקבע על ידי (מנהל המערכת)		מראש, ולהפריד את התוויות מהמסלולים 2. להעלות את המסלולים למסד 3. בחירת כל המסלולים (X) מהמסד 4. הרצת למידה 5. השוואת התוויות שהתקבלו מהלמידה מול התוויות	מאומנת מראש	ולמידה עליהם	
2	אחסון מסלול שעבר למידה במסד עם הסיווג שקיבל כתוצאה מהלמידה	התאמה מלאה בין הסיווג הידוע לבין הסיווג במסד, בין המסלול המקורי לבין המסלול במסד	מסלול כלשהו שידוע לנו הסיווג שלו מראש	1. המכונה הלומדת תקבל מסלול כלשהו ותסווג אותו 2. נבדוק שהסיווג שהתקבל תואם לתווית הידועה מראש - אחרת נדחה 3. לנתק את הקשר למסד ולפתוח מחדש (אם השרת יהיה REST לא יהיה צורך לנתק את הקשר) 4. נמשוך את המסלול מחדש מהמסד ונוודא שהתווית שבמסד נכונה ושאינן שגיאות באחסון המסלול	המערכת צריכה להיות מאומנת מראש		

#### בדיקות מערכת

מס'	תיאור	תנאי קדם	צעדים	דאטה	צפי	תוצאה	עבר/נכשל
1	דימוי זיהוי מסלול אמיתי	המערכת צריכה להיות במצב מנוחה, כמוגדר <a href="#">בתהליכים ומצבי ומערכת</a> . צריכות להיות משקולות בשרת שמעודכנות לאותו פורמט של הדאטה.	1. הרצת מודול חיצוני שידמה חיישנים שכותבים בזמן אמת מסלול (המסלול והסיווג ידועים לנו מראש) 2. קבלת חלק מהנתונים במחשב משתמש וקריאה לשרת לבקשת המשקולות המעודכנות 3. הרצת הסיווג 4. בדיקת התוצאות שאכן יצאו דומות או זהות.	מסלול מופרד מהסיווג שלו.	קבלת $X \cdot 0.8$ קבצים בתיקיה train, $X \cdot 0.2$ קבצים בתיקיה test, כל קובץ עם Y שורות		



6. נספחים

א. טבלת סיכונים

Name	Description / Impact	Prob (%)	Impact (L, M, H, VH)	Risk Code	Mitigation Plan / Contingency Plan
Tight Schedules	Due to the college board deciding to begin the second semester earlier, there will be less time to work on the project	90%	H	1	Crunch.
No solution to classification problem	We won't find a suitable model for machine learning to classify plots	10%	VH	1	The project might be cancelled, or frozen over to next year.
COVID-19	A surge in COVID-19 cases might cause me or one of the supervisors to catch the virus	60%	M	2	I can still work from home if my case won't be too severe. If only one supervisor becomes ill, I can use the other, if both get sick, I'll have no feedback, and might have to rely on Assaf.
New requirements	Any new requirements to come from the client side	30%	H	3	Deviations from schedule.

ב. טבלת דרישות

❖ דרישות פונקציונליות

מס' דרישה	תיאור
1	המערכת שמקבלת את המסלולים צריכה להתאים למסלולים עם X נקודות (ייקבע על ידי הלקוח), ולדעת להתאים מסלולים עם יותר או פחות נקודות
2	מסד נתונים לאגירת מסלול הגוף

❖ דרישות נוספות ולא פונקציונליות

○ מהירות (זמן): המערכת צריכה לעלות תוך מס' שניות (לא יותר מ-5)

מרגע זיהוי גוף ע"י חיישנים חיצוניים

○ זמינות: רק בעת הצורך, המערכת לא תפעל תמיד

ג. תכנון הפרויקט

04.10.21	<<פגישת הכרות עם הלקוח >>
18.10.21	תחילת למידה- מערכות מבוזרות לאחסון נתונים
01.11.21	המשך למידה
15.11.21	כתיבת דוח הצעת פרוייקט, תחילת תיכון אחסון המערכת
29.11.21	סיום תיכון ומעבר לעבודה על תיכון מערכת הסיווג
13.12.21	כתיבת מחולל מסלולים, בדיקת פורמטים לאחסון הנתונים
27.12.21	החלטה על הפורמט, בניית נתונים ללמידה עתידית
10.01.22	חקירת סוגי הלמידה למציאת למידה מתאימה
24.01.22	המשך חקירה
07.02.22	הגעה למסקנות ותחילת כתיבה של תוכנית עבודה למערכת

ד. סקר שוק

1. מסדי נתונים

בדיקת מערכות לאגירת מסלולים במערכות אחרות, לדוג' מערכות רכבים אוטונומיות, מטוסים וכו'. לאחר התייעצות עם המנחה הוסכם כי שנינו מעדיף לעבוד עם מסד נתונים לא-יחסי, כלומר מסוג NoSQL. הבסיסים אותם אני בודק למען הפרויקט הם MongoDB, Cassandra ו-HBase.

a. Cassandra

בסיס נתונים מבוזר הכתוב בשפת ג'אווה, מיועד לטיפול בכמות נתונים גדולה העולה על מאות טרה-בייטים.

הבסיס אינו משתמש במערכת Master-Slave אלא כל מחשב בצומת יוכל לפנות לכל צומת אחר. בנוסף, כל צומת מכיל שכפולים של מס' צמתים אחרים. המסד מבוסס שיטת "אחסון עמודה רחבה", כלומר משתמש בטבלאות עם שורות ועמודות, אך בניגוד למסדי SQL מאפשר ששמות ופורמטי העמודות ישתנו משורה לשורה.

חסרונות מערכת זו כוללים חוסר עקביות בעת שחזור מידע לצומת שנפלה, זמן קריאה ארוך מהשרת במידה והמפתח אינו ידוע למשתמש (לדוג' חיפוש של דוגמא ספציפית), וחוסר של תיעוד רשמי מאת הארגון המפתח, קרן Apache. שימושים ראשיים למסד כוללים תוכנות מסרים מיידיים, מסחר אלקטרוני ועיבוד מידע מחיישנים בזמן אמת.

#### **HBase** .b

בסיס נתונים הכתוב גם הוא בשפת ג'אווה, מבוסס על מערכת Bigtable של חברת גוגל, ומשתמש במספר תכונות שלו לרבות ביצוע פעולות על-גבי הזכרון בזמן אמת (לעומת גישה למידע הנמצא על גבי מחשבים רבים, פעולה הלוקחת הרבה יותר זמן), כיווץ מידע במוסד ושימוש במבני נתונים מבוססי Bloom Filter, אשר יכולים לקצר את זמן ההמתנה למשאבים מהדיסקים באופן משמעותי. בדומה ל-Cassandra, גם HBase מבוסס שיטת "אחסון עמודה רחבה". אחת מחוזקות המסד היא השימוש המובנה שלו ב-HDFS, מערכת הקבצים המבוזרת של Hadoop. המסד משתמש במערכת הקבצים HDFS של Hadoop, כך שאין צורך להשתמש במערכת MapReduce. חסרונות עיקריים של המסד כוללים שימוש במודל Master/Slave, דבר התורם לסיכוי לנקודת כשלון יחידה, אין ל-HBase שפת שאילתות מובנית (צריך להשתמש בממשק חיצוני, לדוג' Apache Hive), ותלות-יתר גבוהה בטכנולוגיות המובנות של Hadoop. שימושים שכיחים ל-HBase כוללים אנליטיקת לוגים מקוונת ואתרים מרובי תוכן (כגון נטפליקס, פינטרסט, פייסבוק).

#### **MongoDB** .c

בסיס נתונים הכתוב בשפות JS, C++ ופייתון, מבוסס שיטת אחסון מסמכים, המציע גרסת קהילה וגרסת עסקים הכוללת גישה מהירה וקלה לנתונים על גבי הרשת, ביקורת נתונים מובנית, אבטחת מידע והצפנה. MongoDB מאחסן את הנתונים על גבי המערכת כמסמכי JSON בינאריים, שיטה המאפשרת גישה מהירה למידע. גרסת העסקים גם כוללת זמינות גבוהה ע"י שכפול המידע ליצירת עודפות וריבוי צמתים לאחסון. חסרונות של המסד כוללות חוסר תמיכה רשמי לגרסת הקהילה (לדוג' תיקוני טלאי צריכים להתבצע באופן ידני), וגרסאות סקלביליות גוזלות יותר משאבי זכרון.

2. אלגוריתמי למידה  
במסגרת הפרויקט אסקור חלק מאלגוריתמי הלמידה האפשריים שיוכלו לשמש אותי  
במסגרת הפרויקט.

a. למידה מונחית

ענף הלמידה הפופולרי ביותר בתחום למידת המכונה, מאפשר פיתוח של מכונה הלומדת על בסיס דוגמאות שהתשובה אליהן ידועה. דומה מאוד לצורת הלימוד של ילד עם כרטיסיות; בהינתן דוגמאות מוכנות עם תוויות ידועות מראש, המכונה תקבל זוגות של דוגמא ותווית, ובהינתן מענה לגבי האם החיזוי של המכונה נכון או לא, היא תוכל 'לשפר' ולחזות יותר טוב בכל איטרציה עתידית ע"י ניתוח הקשר בין תכונות שבתוך הדוגמא לבין התווית. אלגוריתמים ידועים בתחום הינם רשתות נוירונים, אלגוריתם שכן-קרוב ועצי החלטה.

שימושים ידועים של הענף הינם בלמידת תווי פנים וזיהוי ספאם בדוא"ל.

יתרונות בשל שימוש באלגוריתמים מהענף:

- i. מאפשר חיזוי איכותי של מטרות מספריות כגון מס' בתוך תחום.
- ii. שימושי בבעיות סיווג.
- iii. לא צריך להשאיר את הנתונים מהדוגמאות לאחר סיום הלמידה עליהן, מאפשר חיסכון במקום.

חסרונות בשל שימוש באלגוריתמים מהענף:

- i. איכות המכונה תלויה באיכות הדוגמאות, דבר הדורש בחירת דוגמאות איכותיות יותר ולעיתים אף סינון ידני של דוגמאות.
- ii. במקרה של בעיות סיווג, במידה ונכניס למכונה נתונים שלא שייכים לאף מחלקה שנילמדה קודם לכן, ייתכן מאוד כי נקבל סיווג שגוי.
- iii. דורש זמני חישובים ארוכים יותר.

b. למידה בלתי-מונחית

ענף נוסף בתחום אשר בדומה לענף הלמידה המונחית, בו המכונה לומדת מדוגמאות ומנסה לנתח ולהפריד לקבוצות, אך בניגוד ללמידת מכונה, אינה מקבלת תוויות ולכן אינה יודעת האם היא צודקת או לא. לדוגמא, להפריד סוגי פירות שונים לפי הצבע והגודל של הפרי. המכונה תנסה להפריד לקבוצות את הנתונים לפי תכונות הקיימות בכל קבוצת הדוגמאות.

שימוש ידוע של הענף היא המלצות על סרטים, לדוג' Netflix וב-YouTube.

יתרונות בשל שימוש באלגוריתמים מהענף:

- i. מכונה יכולה לזהות דפוסים שלא בהכרח יהיה קל לזהות באופן חזותי.
- ii. למכונות בענף יש פחות סיבוכיות ביחס למכונות בענף הלמידה המונחית, ולכן יותר קל לפתח אותן.
- iii. הרבה יותר להשיג נתונים ללא תוויות.

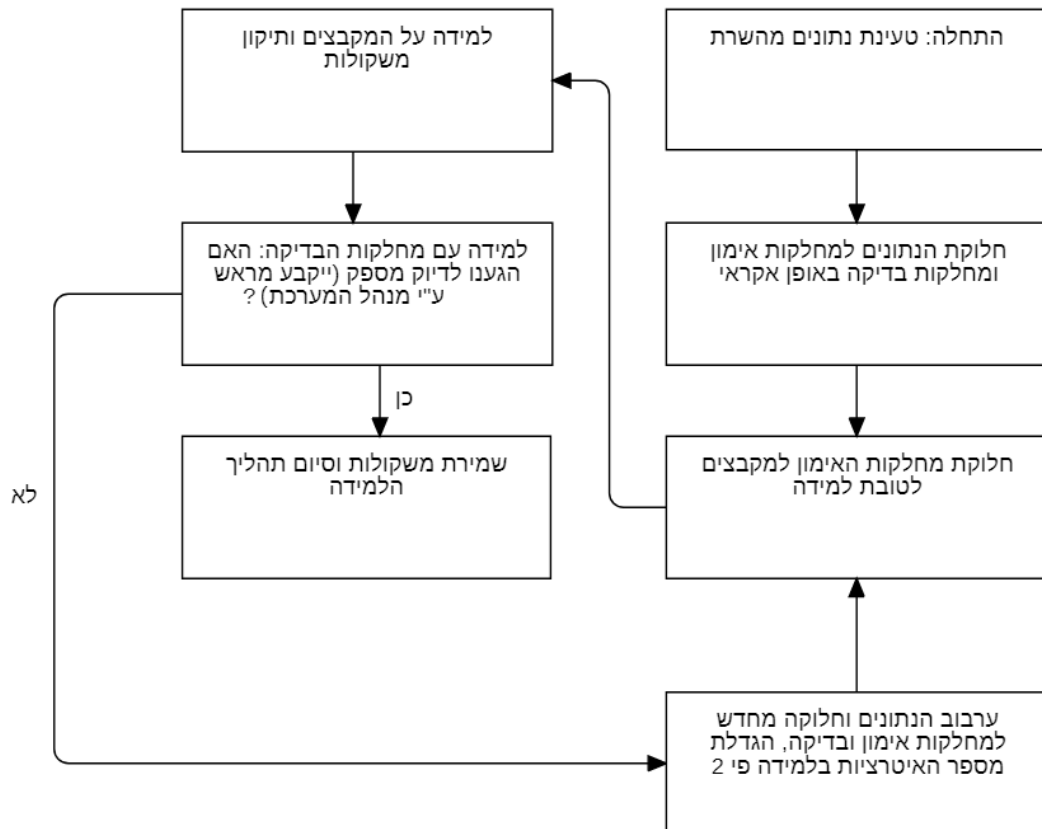
חסרונות בשל שימוש באלגוריתמים מהענף:

- i. לעיתים ידרוש התערבות אנושית כדי להבין את הדפוסים שהמכונה הצליחה לקשר בין הדוגמאות השונות, ולכן יותר יקר.
- ii. רמת הדיוק של מכונות בענף נמוכה יותר ביחס למכונות בענף הלמידה המונחית, לדוג' בחלוקה של דוגמאות לקבוצות לעומת סיווג הדוגמאות באופן ישיר.

ה. תרשימים

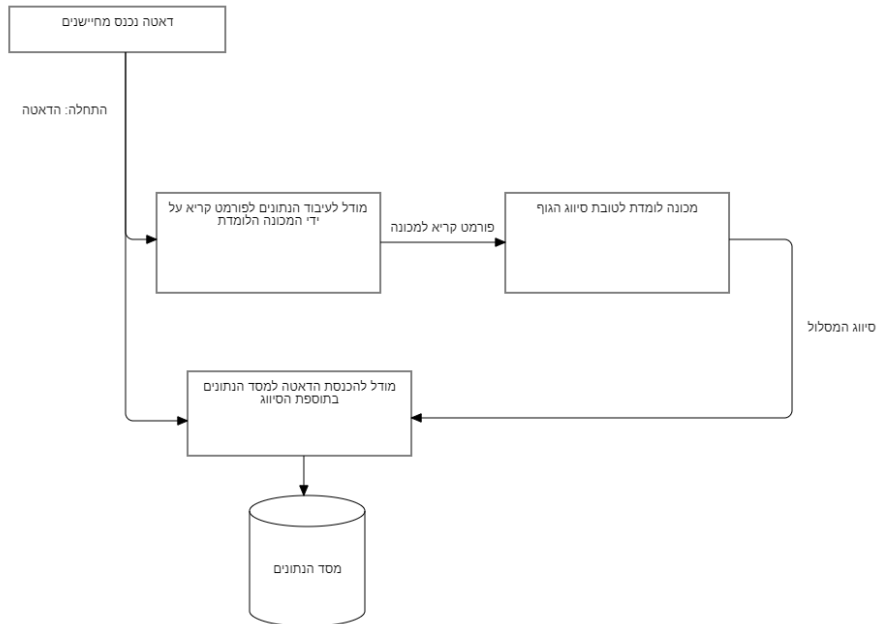
1. תהליך הלמידה

תיאור תהליך הלמידה של המערכת; יתבצע בהפעלה הראשונית ואחת לתקופה התיקבע מראש ע"י מנהל המערכת



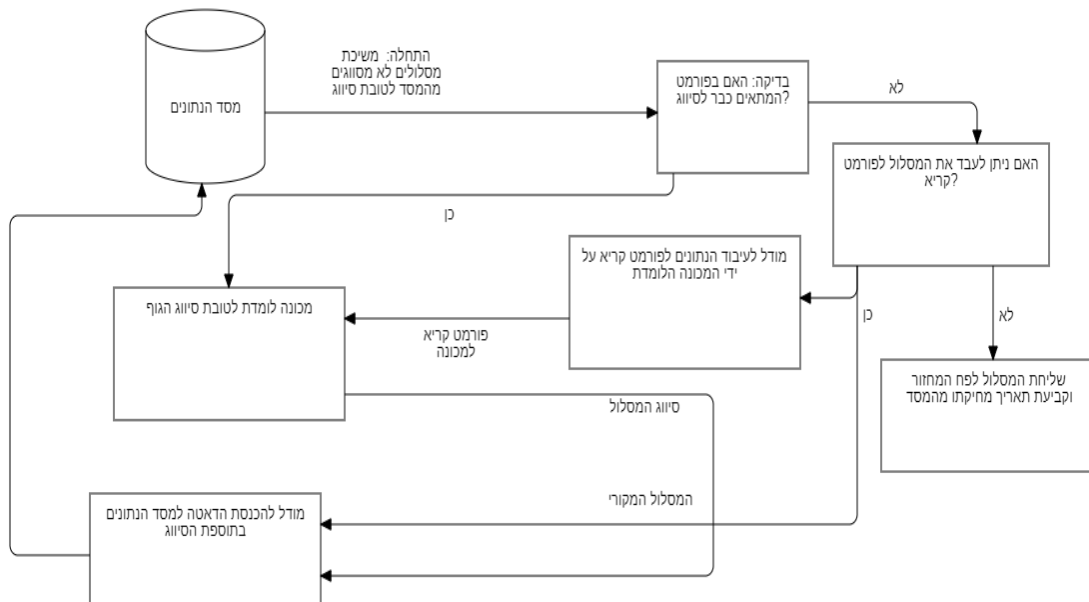
## 2. סיווג מחיישנים

תהליך הסיווג של מסלול חדש שנקלט ע"י חיישנים חיצוניים



## 3. סיווג ממסד

תהליך הסיווג של מסלולים ממסד הנתונים שעוד לא סווגו



ו. רשימת ספרות \ ביבליוגרפיה  
עבור סקר השוק:

<https://logz.io/blog/nosql-database-comparison/>  
<https://db-engines.com/en/system/Cassandra%3BHBBase%3BMongoDB>  
<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0045-4>

מחקרים דומים, עבור הפרק סקירת עבודות דומות:

Bosson C, Nikoleris T (2018 January) Supervised Learning Applied to Air Traffic Trajectory Classification, **2018 AIAA Information Systems-AIAA Infotech @ Aerospace**  
<https://ntrs.nasa.gov/api/citations/20180000801/downloads/20180000801.pdf>

M.K. Sakyi-Gyinae (2019 January) A Machine Learning Approach to Evaluating Aircraft Deviations from Planned Routes, **Transportation Research Part C Emerging Technologies**  
<https://repository.tudelft.nl/islandora/object/uuid:274b4386-539a-4193-80e9-f120c8d4832e/datastream/OBJ/download>