**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 | <https://github.com/EvjaG/FinalProject> |
| 2 | קישור ליומן | <https://mail313575.monday.com/boards/2070942680/> |
| 3 | קישור לסרטון דוח אלפא | להקלטת מסך המחשב ודיבור מלווה, אפשר להיעזר במדריך והכלים במקור **[6] -**למשל screencast-o-matic, viewedit.com או <https://www.flashbackrecorder.com/express/>  <https://shotcut.org/> או באמצעים של תוכנות תקשורת כמו Google Hangout , Skype (היתרון של חלקם שנותנים שרות אחסון לסרטון ואפילו קוד להצגת הסרטון על גבי דף ההגשה). כמו כן מומלץ להתרשם מסרטוני המסירה ברשימת הפרויקטים משנה שעברה: בקישור המצורף:  [6] הדרכה על הקלטת סרטונים <http://the-openclass.org/core/item/30> וגם עם <https://www.flashbackrecorder.com/> |

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

כלי פיתוח

|  |  |
| --- | --- |
| צד שרת | Python |
| DB | MongoDB |
| VSCode | כללי - Docker |

תוכן עניינים

[0. מבוא 3](#_Toc92222113)

[1. תיאור הבעיה 3](#_Toc92222114)

[2. סקירת עבודות דומות והשוואה 3](#_Toc92222115)

[3. תיאור הפתרון 4](#_Toc92222116)

[דרישות ואפיון הפתרון 5](#_Toc92222117)

[הבעיות מבחינת הנדסת תוכנה 5](#_Toc92222118)

[המשך הפרק תיאור הפתרון 4](#_Toc92222119)

[מהי המערכת 5](#_Toc92222120)

[תהליכים ונתוני המערכת 5](#_Toc92222121)

[תיאור הכלים המשמשים לפתרון 5](#_Toc92222122)

[4. מה עשינו עד כה ? 5](#_Toc92222123)

[**5.** תכנית בדיקות 6](#_Toc92222124)

[1. בדיקות יחידה 6](#_Toc92222125)

[6. נספחים 6](#_Toc92222126)

[א. טבלת סיכונים 6](#_Toc92222127)

[ב. רשימת\טבלת דרישות 6](#_Toc92222128)

[ג. תכנון הפרויקט – ברזולוציה של שבועיים. 7](#_Toc92222129)

[ד. רשימת ספרות \ ביבליוגרפיה 8](#_Toc92222130)

# מבוא

כיום מוערך שיש בעולם כ-47,840 כלי טיס צבאיים (WMMDA) 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.

# תיאור הבעיה

הבעיה – המצב הקיים כיום – הוא שX.

# סקירת עבודות דומות והשוואה

לא הצלחתי לקבל מאף חברה שיצרתי איתה קשר מידע לגבי המערכות בהן הן נעזרות לצורך קטלוג וזיהוי, ולכן ביצעתי בהנחיית ראובן סקר שוק לגבי סוגי מסדי נתונים שיכולים להתאים לכמות גדולה של נתונים. השוואת מסדי הנתונים השונים נמצאת בנספחים. בנוסף , אני מתכנן לעשות השוואה

# 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.

## הבעיות מבחינת הנדסת תוכנה

* הקמת מערכת מבוזרת שתוכל לאגור ולקטלג את המידע, ולמשוך אותו בשלב יותר מתקדם.
* בניית מערכת שתמדוד את ביצועי האלגוריתמים ותשווה ביניהם.
* בניית מחולל מסלולים לצורכי הזנת נתונים ולמידה מוקדמת.
* 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.

ממשק הפעלה למערכת.

הפתרון אליו הגענו מכיל מס' רכיבים, ובהם:

* מחולל מסלולים, לצורך יצירת מסלולים בסיסיים עליהם נלמד בהמשך (היות והמסלולים הקיימים במערכות של החיל הם חסויים) ומערכת הכותבת מסלולים אלו לקבצי תיעוד וקבצים ללמידה.
* מערכת למידה – יצריך ככל הנראה רשת נוירונים, לא בטוח עדיין אם CNN או ANN #TODO
* מסד נתונים – נוכל לאחסן בו את כל המסלולים והמידע שקשור ללמידת המכונה לצורך שליפה מהירה במידת הצורך (לדוג' לזיהוי אובייקט הנע במהירות במרחב)
* רק אחר כך (אפשר בתתי פרק) פירוט על אופן הפתרון עצמו - הטכנולוגיות שמשתמשים בהן, חלוקה למודולים, High Level Design, תרשים זרימה אם רלוונטי, פירוט של אלגוריתמים מוכרים שנעשה בהם שימוש (עם חלוקת קרדיט והפניה לביבליוגרפיה), פירוט אלגוריתמים יעודיים שהמצאתם\המצאתי\המצאנו (בלי חלוקת קרדיט) וכו'.
* מצופה כבר outline של הפתרון ממעוף הציפור, אבל כמובן אין עדיין חובה להיכנס לפרטים - ידוע לנו שחלק מהפרויקט זה לגלות איך לפתור את הבעיות, ושכרגע הדברים עדיין פתוחים. אתם יכולים להתייחס לחוסר הודאות כרגע במפורש: "החלק הזה ייפתר על ידי שימוש ב-X או ב-Y, כתלות במי יתברר כיותר מהיר".

# המשך הפרק תיאור הפתרון

במקרה שיש הרבה איורים יש להפנות לנספחים מטה (ראו שם).

פרק זה יכלול את כל התרשימים אשר למדתם -- תרשימי שימוש, ארכיטקטורת המערכת, אלגוריתמים , תיאור הקוד ותיכון (תרשימי הפצה ומודולים עיקריים, שפות וכלים מתוכננים לשימוש כגון  מערכות הפעלה, שימוש ברכיבים קיימים, סביבות, כלי בדיקה). תרשימי תיכון כגון דיאגרמת רכיבים \ הפצה (UML), דיאגרמת ישויות טבלאות במסד נתונים

מעבר לזה –המרכיבים העיקריים לתיאור תת-פרויקט זה (כמו שלמדתם בקורס בהנדסת תוכנה).

1. Use-cases
2. Requirements for relevant use-case
3. Design (תיכון).

|  |
| --- |
| Use Cases כאמור מיצגים את צד הלקוח  Requirements מייצגות את צד המערכת  לעיתים הלקוח יודע מה הוא רוצה , לדוגמא - תמיכה בIOS גרסאות 9 ומעלה , או לא רוצה בכלל תמיכה הניידים אלא רק במחשב נייח ומערכת הפעלה Win 10, אלו דרישות מערכת    לכן לא לכל Use Cases יש דרישות ישירות  ההפרדה, כפי שלמדתם היא:   1. Use Cases לתואר הפעולה שהלקוח יעשה -- ולרב הפעולות האלו בסוף יהיו מסכים או חלקים ממסך 2. דרישות המערכת  -- שיותר כוללות מערכות הפעלה , סוגי ניידים ( אם בכלל יש תמיכה במובייל), גרסאות תוכנה , חיבורים  מסד נתונים ( בדגש של איזה ואיזו גרסאות )    צריך לוודא שכל דרישה היא גם ברת בדיקה |
|  |

## מהי המערכת

<<התרשים הבא (מספר ..) מתאר את ארכיטקטורת המערכת ומורכב מ ... >>

## תהליכים ונתוני המערכת

<<המערכת תכלול 2 מצבים של עבודה:>>

## תיאור הכלים המשמשים לפתרון

<<בפרויקט השתמשתי ...>>

# מה עשינו עד כה ?

מכיוון שהדוח כבר במבנה דוח סופי, אבל זה עדיין לא דוח סופי. קשה להבין מה נעשה ומה עוד צריך להיעשות. ולכן כאן לרשום במפורש. ראו דוגמא:

Text

Description automatically generated

מה בוצע עד כה ומסקנות מעניינות שעלו מביצוע הפרויקט. **מסקנות -** לקחים מעניינים מהפרויקט עד כה (אם יש) ודברים שחשוב לצייין להמשך.

* יצירת מחולל מסלולים, כולל כתיבה לקבצי PNG וCSV
* יצירת מודול לקריאת הCSV ופרסור (לדו’' מציאת המהירות בכיוון X בזמן T)
* בחירת מסד נתונים לעבודה ראשונית
* יצירת מכונה לומדת לסיווג המסלולים
* ביצוע ניסויים ראשוניים ללמידה, ללא בדיקה
* בדיקת המסלולים במחלקת test
* כתיבת מודול לאחסון ושליפת המסלולים ממסד הנתונים

# תכנית בדיקות

על המערכת יבוצעו שלוש סוגי בדיקות, אותן אתאר בפירוט בפרק הקרוב:

## בדיקות יחידה

* המודולים לקריאה וכתיבה של קבצי CSV
* נכונות/רמת דיוק של המערכת הלומדת
* מערכת שתדאג לאחסן את המסלולים במסד נתונים בצורה נכונה ולשלוף אותם ממנו במידת הצורך
* מודול שיוכל להתאים דאטה שייכנס מתוך מערכת חיצונית (לדוג' מחיישנים) לפורמט המוסכם

בהמשך הדרך ייתכן כי נחליט שנכון לבנות מערכת קבצים מבוזרת (כדוג' Hadoop). אם כן,

נצטרך לבנות מערכת להזנת המסלולים לתוך המסד שישב על גבי מערכת הקבצים המבוזרת  
בצורה נוחה ומהירה, ולבדוק גם אותו.

## בדיקות אינטגרציה

* שליפה של מסלול חדש ממסד הנתונים ולמידה עליו, ללא מידע חדש מהחיישנים.
* אחסון מסלול שעבר למידה במערכת עם הסיווג שקיבל כתוצאה מהלמידה

## בדיקות מערכת

איזה בדיקות אפשר לעשות ? מה הם הבדיקות ? – חזרו לעיין בשיעור המתאים מהקורס הנדסת תוכנה

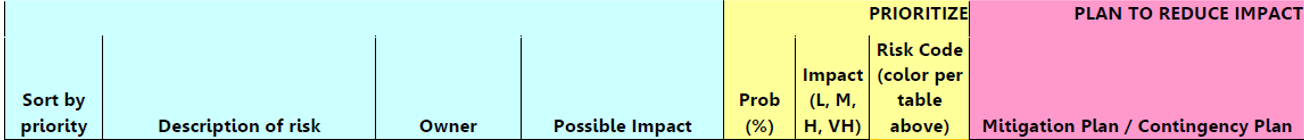
# נספחים

ספרות, תרשימים נוספים, תכנון הפרויקט, טבלת ניהול סיכונים, טבלת דרישות (URD),

**במקרה שלא היה מקום פרק "תיאור הפתרון" דאגרמות עם חשיבות משנית אפשר לשים בנספח במסכים (אם לא למעלה)**

## טבלת סיכונים

עדכון מצב הסיכונים– ע"פ המבנה הבאה



* עדכון רשימת המשימות והסיכונים – יש להשתמש בtemplate סיכונים מהקרוס "הנדסת תוכנה"

## רשימת\טבלת דרישות

- עדכון אם נוספו/בוטלו/השתנתה עדיפות וכדו'

פורמט טבלת הדרישות יהיה לפי המקובל בארגון. להלן דוגמא:

**טבלת דרישות (User Requirement Document)**

* **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 |  |
| 3 |  |
| 4 |  |
| 5 | מסד נתונים לאגירת מסלול הגוף |

* **דרישות נוספות ולא פונקציונאליות**
  + **מהירות (זמן): המערכת צריכה לעלות תוך מס' שניות (לא יותר מ-5) מרגע זיהוי גוף ע"י חיישנים חיצוניים**
  + **עומס:**
  + **זמינות: רק בעת הצורך, המערכת לא תפעל תמיד**
  + **קיבול (צריכת זיכרון / CPU):**
  + **ניידות:**
  + **שימושיות:**
  + **תחזוקה:**
  + **אילוצי פלטפורמת מימוש:**

## תכנון הפרויקט – ברזולוציה של שבועיים.

|  |  |
| --- | --- |
| 4.10. | <<פגישת הכרות עם הלקוח >> |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| 22.2.13 | <<סיום פיתוח שרת>> |
| 22.3.13 |  |

## רשימת ספרות \ ביבליוגרפיה