**המחלקה להנדסת תוכנה**

**פרויקט גמר – תשפ"ב**

אגירה וסיווג של מסלולי גופים פיזיקליים באמצעות למידת מכונה  
Storage and Classification of Trajectories via Machine Learning Methods

**4 עמודים**

**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 | יומן | ?? |

מידע נוסף (מחק את המיותר)

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

מבוא

מעבדת התעשייה הביטחונית אוספת במשך שנים רבות מסלולי גופים, כעת מעוניינים באמצעים של למידת מכונה לנצל נתונים אלו, על מנת להשוות לנתונים בזמן אמת ולחזות שינויים שיכולים לקרות לגוף בתנועה.

דרישות ואפיון הבעיה

## דרישות ואפיון הבעיה

הקמת מערכת הכוללת בסיס נתונים ומערכת לאגירת נתונים בזמן-אמת מגופים פיסיקליים הנעים במרחב, הכולל תיעוד מסלולם ותכונותיהם. המאגר יאחסן ויקטלג את הנתונים בעזרת הכלים של למידת מכונה ויאפשר חיזוי בזמן אמת. 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. בנוסף, בניית מערכת שתמדוד את ביצועי האלגוריתמים ותשווה ביניהם. בניית מחולל מסלולים לצורכי הזנת נתונים ולמידה מוקדמת. ממשק הפעלה למערכת.

תיאור הפתרון

## **רשימת\טבלת דרישות**

|  |  |
| --- | --- |
| דרישה | צורך |
| זמינות מערכת | בעת הצורך, מערכת לא תפעל תמיד |
| מודולריות | גבוהה, במידה ויהיה צורך להחליף חלק אחד במערכת באחד, שלא יצרוך יותר מדי עבודה |
| מערכת הזנת נתונים | צריכה להיות זמינה בכל עת פעילות המערכת, ותאפשר להכניס נתונים לתוך המסד בזמן אמת |
| למידת מכונה | המערכת צריכה לאפשר למידה של מסלול חדש מול מסלולים קיימים ע"מ לחזות שינויים אפשריים |
| השוואת ביצועים | רצוי שהמערכת תכלול מס' אלגוריתמים להשוואת הביצועים של מסלולי טיסה שונים. |
| מערכת הפעלה | לא הוחלט עדיין |
| דרישות חומרה | לא הוחלט עדיין |

## (UML)

Diagram

Description automatically generated

על הלקוח

ניתן לקרוא על הגוף בו פועלת המעבדה <https://www.iaf.org.il/4968-36146-he/IAF.aspx>

נספחים

## **תכנון הפרויקט**

|  |  |
| --- | --- |
| 30.11 | למידת מערכות Hadoop ותכנון מערכת Geohash |
| 15.12 | הקמת מסד נתונים למערכת ומערכת הזנת נתונים |
| 30.12 | תכנון מערכת למידה ומערכת אוטומטית להעברת מידע לשרת בעזרת MapReduce |
| 30.01.22 | סיום בניית מודל אלפה ראשוני |
|  | יצירת מחולל מסלולים לבדיקה |
| 28.02.22 | בניית מערך/סקריפט הפצה כללי לכלל מערכת Hadoop למספר מחשבים רב |
| 30.04.22 | הטמעת אלגוריתמים לחישוב ביצועים |
| 01.06.22 | הרצות וביצועי מערכת |
| 30.06.22 | סיום פרוייקט? |

## **סקר שוק**

בדיקת מערכות לאגירת מסלולים במערכות אחרות, לדוג' מערכות רכבים אוטונומיות, מטוסים וכו'.

בדיקת סוגי מסדי הנתונים השונים

לאחר התייעצות עם המנחה הוסכם כי שנינו מעדיף לעבוד עם מסד נתונים לא-יחסי, כלומר מסוג NoSQL. הבסיסים אותם אני בודק למען הפרוייקט הם MongoDB, Cassandra ו-HBase.

**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 הפנות 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 מסדי SQL מאפשר ששמות ופורמטי העמודות ישתנו משורה לשורה.   
*חסרונות* הבסיס 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 מיידים, מסחר אלקטרוני ועיבוד מידע מחיישנים בזמן אמת.

**HBase**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**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.

ספרות:  
[**https://logz.io/blog/nosql-database-comparison/**](https://logz.io/blog/nosql-database-comparison/)[**https://www.mongodb.com/compare/mongodb-hbase**](https://www.mongodb.com/compare/mongodb-hbase)[**https://db-engines.com/en/system/Cassandra%3BHBase%3BMongoDB**](https://db-engines.com/en/system/Cassandra%3BHBase%3BMongoDB)[**https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0045-4**](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0045-4)