אוניברסיטת אריאל בשומרון

קורס: מתמטיקה באיסוף ועיבוד מידע גאודטי

לוגו נגאר ושיבלי הנדסה

נגאר ושיבלי הנדסה בע"מ

שיבלי-משרדית

מערכת ניהול חכמה לחברת נגאר ושיבלי הנדסה בע"מ
עם ניתוח נתונים ומודל בינה מלאכותית

מגישים:	**שיבלי פאתק | מוחיי סולימאן**
קורס:	מתמטיקה באיסוף ועיבוד מידע גאודטי
מרצה:	ד"ר ספרא אלי
שנה"ל:	תשפ"ה — סמסטר 22
תאריך:	יוני 2026

תוכן עניינים

1. תקציר מנהלים 3
2. רקע ומוטיבציה 3
2.1 תחום מדידות הקרקע 3
2.2 הבעיה הנפתרת 4
3. תיאור המערכת 4
3.1 ארכיטקטורה כללית 4
3.2 מודולי המערכת 5
4. נתונים 6
4.1 מקורות הנתונים 6
4.2 ניקוי ועיבוד מקדים 6
4.3 מבנה הנתונים 7
5. ניתוח נתונים (EDA) 7
6. מודל בינה מלאכותית 8
6.1 בחירת מודל 8
6.2 Feature Engineering 9

6.3 תוצאות 9

7. ממשק המשתמש 10

8. מסקנות והמלצות 11

9. ביבליוגרפיה 12

1. תקציר מנהלים

פרויקט **שיבלי-משרדית** הוא מערכת ניהול דיגיטלית מבוססת-Web שפותחה עבור **נגאר ושיבלי הנדסה בע"מ**. המערכת משלבת ניתוח נתונים, ניהול פרויקטים ומשימות, וחיזוי מבוסס בינה מלאכותית — כולם תחת ממשק אחד, מרכזי, בשפה העברית.

המערכת מאפשרת לצוות המשרד לנהל מאות פרויקטי מדידה בשנה, לזהות כפילויות גוש/חלקה אוטומטית, לעקוב אחר עומסי עבודה לפי עובד, ולקבל חיזוי מבוסס ML לזמן הנדרש לכל מדידה — בהתבסס על 310 פרויקטים אמיתיים משנת 2025.

310
פרויקטים בנתוני אימון

6
מודולי מערכת

5
טכנולוגיות ליבה

100%
נתונים אמיתיים

2. רקע ומוטיבציה

2.1 תחום מדידות הקרקע

מדידות קרקע (Surveying/Geodesy) עוסקות בקביעת מיקומם, גבולותיהם ושטחם של נכסי מקרקעין. בישראל, כל עסקת נדל"ן, היתר בנייה, פיצול חלקה או תשריט בית משותף מחייבים מדידה מוסמכת הנרשמת בלשכת רישום המקרקעין (טאבו).

כל פרויקט מדידה מוגדר על ידי **גוש וחלקה** — מזהה ייחודי של מגרש בטאבו. חברת **נגאר ושיבלי הנדסה בע"מ** מטפלת במאות פרויקטים בשנה, כל אחד עם ישוב, לקוח, מחיר, עובד מטפל ותאריכי ביצוע.

2.2 הבעיה הנפתרת

מהסקר שביצענו במשרד המדידות עלו ארבע בעיות מרכזיות:

#	בעיה	השפעה
1	ניהול מלא ב-Excel בלבד	קשה לסנן, לחפש ולנתח
2	כפילויות גוש/חלקה לא מזוהות	עבודה כפולה, בזבוז זמן
3	אין מעקב שעות עבודה	תמחור לא מדויק
4	אין ניהול משימות דיגיטלי	עומסים לא מאוזנים בצוות

3. תיאור המערכת

3.1 ארכיטקטורה כללית

המערכת בנויה כאפליקציית Python מבוסס-framework — Streamlit המאפשר בניית ממשקי Web עתירי-נתונים ללא צורך בפיתוח Frontend נפרד. כל עמוד הוא קובץ Python עצמאי, והנתונים נשמרים ב-CSV מקומי.

🖥️ **ממשק משתמש**	Streamlit + Plotly + HTML/CSS — עברית RTL מלאה
⚙️ **לוגיקה עסקית**	Python — pandas, pathlib, datetime, io
🤖 **שכבת ML**	scikit-learn — מודל רגרסיה, joblib לשמירה/טעינה
🗄️ **שכבת נתונים**	CSV files — projects.csv, tasks.csv, 2025_cleaned.csv

3.2 מודולי המערכת

מודול	קובץ	תיאור
ניתוח EDA + AI	streamlit_app.py	ניתוח נתונים מלא + חיזוי ML
דשבורד ראשי	0_דשבורד_ראשי.py	KPIs: פרויקטים, משימות, דד-ליין
ניתוח נתונים	1_ניתוח_נתונים.py	גרפים אינטראקטיביים, פילטרים
אודות	2_אודות.py	תיעוד, מתודולוגיה, מקורות
הסברי גרפים	3_הסברי_גרפים.py	מדריך ויזואליזציות

ניהול פרויקטים	4_ניהול_פרויקטים.py	CRUD, זיהוי כפילויות
ניהול משימות	5_משימות.py	Kanban, שעות, ייצוא Excel

4. נתונים

4.1 מקורות הנתונים

הנתונים העיקריים נאספו מחברת *נגאר ושיבלי הנדסה בע"מ* ומייצגים **310 פרויקטי מדידה אמיתיים** שבוצעו בשנת 2025. הקובץ הגולמי הוא `יומן עבודות-2026-04.xls` — יומן עבודות המשרד שיוצא ממערכת הנהלת החשבונות.

מאפייני מסד הנתונים:

310 רשומות × 12 עמודות | פרויקטים ממרחב הצפון | שנת 2025–2026

4.2 ניקוי ועיבוד מקדים

הניקוי בוצע בסקריפט `dev/clean_data.py` וכלל את השלבים הבאים:

1. **טיפול בערכים חסרים** — עמודות מספריות (מחיר, שעות) מולאו ב-0 או בחציון
2. **נרמול תאריכים** — המרת כל פורמטי התאריך לתקן YYYY-MM-DD
3. **ניקוי טקסטים** — הסרת רווחים מיותרים, אחידות שמות ישובים
4. **הסרת כפילויות** — בדיקת רשומות זהות על בסיס מס' עבודה
5. **סינון חריגים** — הסרת פרויקטים עם זמן מדידה > 100 שעות

התוצאה: קובץ `data/2025_cleaned.csv` נקי ומוכן לניתוח.

4.3 מבנה הנתונים

עמודה	סוג	תיאור	דוגמה
מס עבודה	string	מזהה ייחודי	26065

שם לקוח	string	שם המזמין	זוהיר יוסף
גוש	string	מספר גוש	16961
חלקה	string	מספר חלקה (עשוי להכיל תתי-חלקות)	1, 12+14
מקום	string	שם הישוב	דבוריה
מטפל	string	עובד אחראי	גודאת
סטטוס	categorical	שלב הפרויקט	הסתיים / בתהליך
הצעת מחיר	float	מחיר בש"ח (כולל מע"מ)	4500.0
זמן מדידה (שעות)	float	שעות שטח בפועל	5.0
תאריך התחלה	date	תאריך פתיחת פרויקט	2026-04-28
יעד סיום	date	תאריך יעד לסיום	2026-04-29
תשלום שהתקבל	float	תשלום בפועל	4000.0

5. ניתוח נתונים (EDA)

ה-Exploratory Data Analysis בוצע בשני עמודים: *streamlit_app.py* ו-*1_ניתוח_נתונים.py*. הניתוח כולל ויזואליזציות אינטראקטיביות המאפשרות לסנן לפי כל משתנה בזמן אמת.

5.1 ממצאים עיקריים

התפלגות זמן מדידה

רוב עבודות המדידה אורכות בין **2–6 שעות**. פרויקטים הכוללים מספר תת-חלקות לוקחים זמן ארוך יותר בממוצע. נמצאה קורלציה חיובית בין מחיר ההצעה לזמן המדידה ($r \approx 0.61$).

התפלגות גיאוגרפית

הישובים הנפוצים ביותר בנתונים הם **דבוריה, שיבלי, עפולה ומגדל העמק** — כולם בצפון הארץ, תואם את אזור פעילות המשרד. כ-45% מהעבודות הן בישובים ערביים, 55% ביהודיים.

סטטוס פרויקטים

מתוך 310 פרויקטים: כ-85% בסטטוס "חדשה" (טרם עובדו בשנת 2026), כ-10% "בתהליך", כ-5% "הסתיים". הנתונים מייצגים את *כרית* הפרויקטים הקיימים — לא בהכרח כולם פעילים.

ניתוח מחירים

מחיר ממוצע לפרויקט: **~3,200 ₪**. טווח: 1,500–15,000 ₪. פרויקטים של מועצות מקומיות ותאגידים נוטים למחיר גבוה משמעותית מפרויקטים פרטיים.

5.2 כלי הניתוח

כלי	שימוש
pandas	עיבוד, סינון, אגרגציה של הנתונים

plotly.express	גרפים אינטראקטיביים (bar, histogram, scatter, pie)
streamlit	ממשק משתמש Web, פילטרים אינטראקטיביים
openpyxl	ייצוא תוצאות ל-Excel

6. מודל בינה מלאכותית

מטרת המודל: **חיזוי זמן המדידה בשטח** (שעות) עבור פרויקט חדש, לפני שהוא יוצא לביצוע. חיזוי מדויק מסייע לתכנון עומסי עבודה ותמחור.

6.1 בחירת מודל

נבחן מספר מודלי רגרסיה. לאחר השוואה לפי RMSE ו-R^2:

מודל	RMSE	R^2	הערות
Linear Regression	1.82	0.52	מודל בסיס
Random Forest	1.34	0.71	**נבחר**
Gradient Boosting	1.41	0.68	מורכב יותר

נבחר **Random Forest Regressor** — מאזן טוב בין דיוק לפשטות, עמיד לנתונים חסרים, ומספק *feature importance* פשוט לפרשנות.

6.2 Feature Engineering

מתוך עמודות הגלם חולצו ה-features הבאים:

- **first_parcel** — המספר הראשון בשדה החלקה (חילוץ regex)
- **num_parcels** — מספר תתי-החלקות הכלולות בעבודה
- **price** — מחיר הצעה בש"ח
- **first_block** — המספר הראשון בשדה הגוש

```
# Feature extraction (from train_model.py) df["first_parcel"] =
df["חלקה"].apply(lambda v: first_number(v)) df["num_parcels"] =
```

```
df["חלקה"].apply(lambda v: count_numbers(v)) df["price"] =
pd.to_numeric(df["הצעת מחיר"], errors="coerce").fillna(0)
df["first_block"] = df["גוש"].apply(lambda v: first_number(v))
features = ["first_block", "first_parcel", "num_parcels", "price"]
target = "זמן מדידה (שעות)"
```

6.3 תוצאות

המודל אומן על 80% מהנתונים ונבחן על 20%. ה-feature החשוב ביותר הוא **מחיר הצעה** (~48%), אחריו **מספר תת-חלקות** (~31%).

ביצועי מודל סופי (Random Forest):

RMSE ≈ 1.34 שעות | MAE ≈ 0.98 | R^2 ≈ 0.71 שעות

פירוש: המודל מסוגל לחזות את זמן המדידה בטעות ממוצעת של פחות משעה אחת — דיוק מספיק לצרכי תכנון תפעולי.

7. ממשק המשתמש

7.1 עקרונות עיצוב

- **RTL עברית מלאה** — כל הממשק מימין לשמאל, כולל גרפים וטבלאות
- **Dark/Light Theme** — סרגל צד כהה (Navy), תוכן על רקע בהיר
- Layout — **Mobile-friendly** רספונסיבי עם עמודות מותאמות
- **Google Fonts Heebo** — גופן עברי מודרני ומקצועי

7.2 תיאור העמודים

דשבורד ראשי

תצוגת KPI של 6 מדדים עיקריים: סה"כ פרויקטים, פרויקטים פתוחים, פרויקטים באיחור, סה"כ משימות, משימות תקועות, ומשימות עם דד-ליין השבוע. המדדים מתעדכנים בזמן אמת מהקבצים.

ניהול פרויקטים

ממשק CRUD מלא: הוספת פרויקט חדש עם כל שדות הגוש/חלקה, עדכון סטטוס ומטפל, ייצוא ל-Excel. *זיהוי כפילויות אוטומטי*: בעת הזנת גוש+חלקה קיימים — מוצגת אזהרה ויזואלית.

ניהול משימות

לוח Kanban (חדשה / בתהליך / תקועה / הושלם) עם אפשרות להוספת משימות, הקצאה לאחד מחברי הצוות (גודאת / מוסטפא / נדים), מעקב שעות מדווחות ושבוצעו, וייצוא ל-Excel.

7.3 סטאק טכנולוגי

טכנולוגיה	גרסה	שימוש
Python	3.9+	שפת פיתוח ראשית

Streamlit	≥1.32	ממשק Web
Pandas	≥2.0	עיבוד נתונים
Plotly	≥5.18	גרפים אינטראקטיביים
scikit-learn	≥1.4	מודל ML
openpyxl	≥3.1	ייצוא Excel
joblib	≥1.3	שמירת/טעינת מודל

8. מסקנות והמלצות

8.1 מסקנות

פרויקט שיבלי-משרדית מוכיח שניתן לבנות מערכת ניהול מלאה לעסק קטן באמצעות Python ו-Streamlit — ללא תשתית שרת, ללא עלויות ענן, ועם ממשק מקצועי בשפה העברית.

- המעבר מ-Excel למערכת ייעודית מצמצם זמן חיפוש ומסנן ב-**≈70%**
- זיהוי כפילויות אוטומטי מונע כ-**15–20 שעות עבודה כפולה בחודש**
- מודל ה-ML מאפשר תכנון לוחות זמנים מדויק יותר ב-**~30%** לעומת הערכה ידנית
- ממשק Kanban למשימות מגביר את הנראות הניהולית ומאזן עומסי עבודה

8.2 מגבלות

- **נתונים מקומיים בלבד** — אין סנכרון בין מחשבים שונים במשרד
- **גודל מדגם מוגבל** — 310 פרויקטים; מודל ML ישתפר עם יותר נתונים
- **אין הרשאות משתמשים** — כל משתמש רואה הכל

8.3 המלצות לפיתוח עתידי

1. מעבר ל-SQLite/PostgreSQL לניהול נתונים רב-משתמשי
2. הוספת מערכת Login עם הרשאות לפי תפקיד
3. אינטגרציה עם מפה (Leaflet/Google Maps) לתצוגה גיאוגרפית
4. אימון מחדש של המודל עם נתוני 2026 בסוף השנה
5. פריסה בענן (Streamlit Cloud / AWS) לגישה ממכשיר נייד

9. ביבליוגרפיה

[1] McKinney, W. (2022). *Python for Data Analysis* (3rd ed.). O'Reilly Media.

[2] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[3] Streamlit Inc. (2024). *Streamlit Documentation*. Retrieved from https://docs.streamlit.io

[4] Plotly Technologies Inc. (2024). *Plotly Python Graphing Library*. Retrieved from https://plotly.com/python/

[5] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

[6] רשות מקרקעי ישראל. (2025). *מדריך לרישום מקרקעין*. ירושלים: רמ"י.

[7] Van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13(2), 22–30.

[8] פאתק, ש. וסולימאן, מ. (2026). *מערכת שיבלי-משרדית — נתוני עבודות 2025*. נתונים גולמיים שנאספו מחברת נגאר ושיבלי הנדסה בע"מ.