

מבוא ל

DATA SCIENCE מדעי הנתונים

A QUICK INTRODUCTION TO

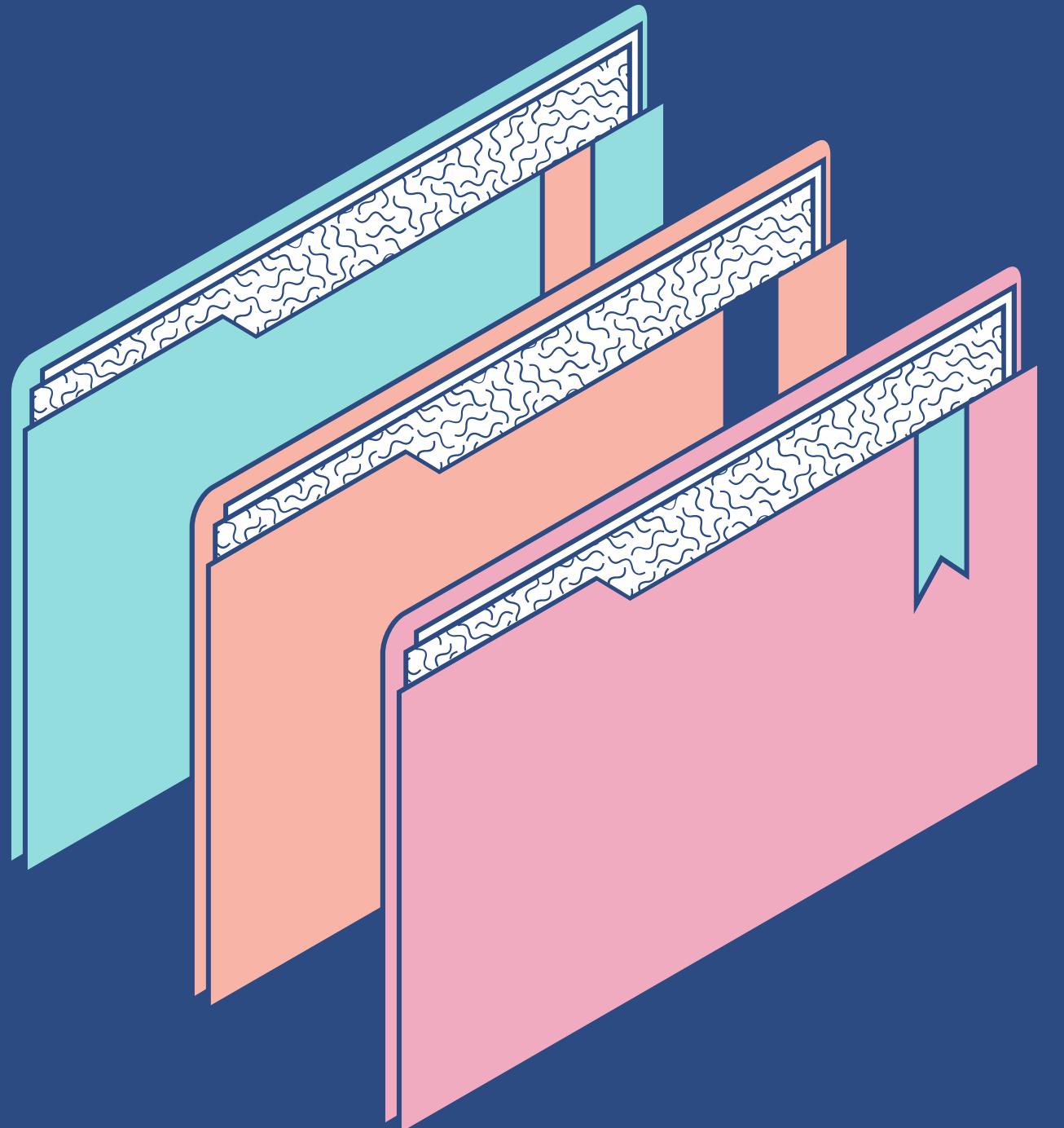
By Sapir Shamay



Agenda

על אילו שאלות ננסה לענות
היום?

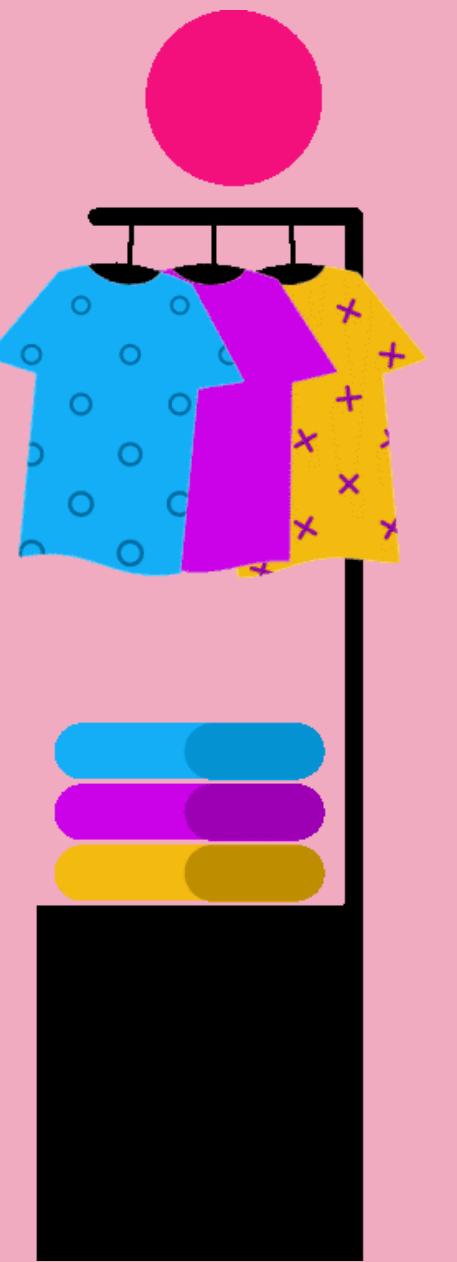
- מהו תחום מדעי הנתוניים?
- ממה שלבים הקיימים בתהליך הפקת הידע?
- מה הקשר של מדעי הנתוניים לבינה מלאכותית?
- ואיך כל זה קשור לשકיעת הטיטאניון?

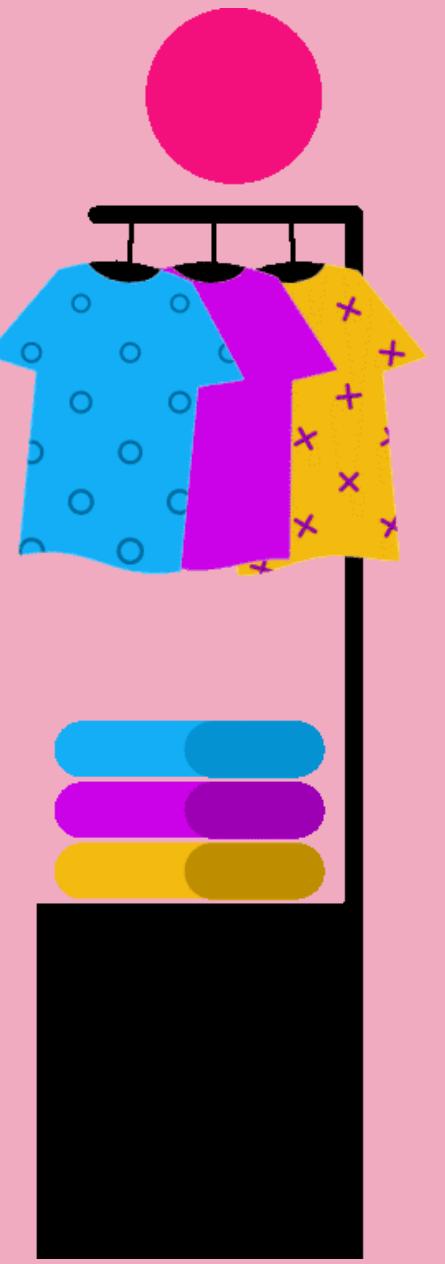


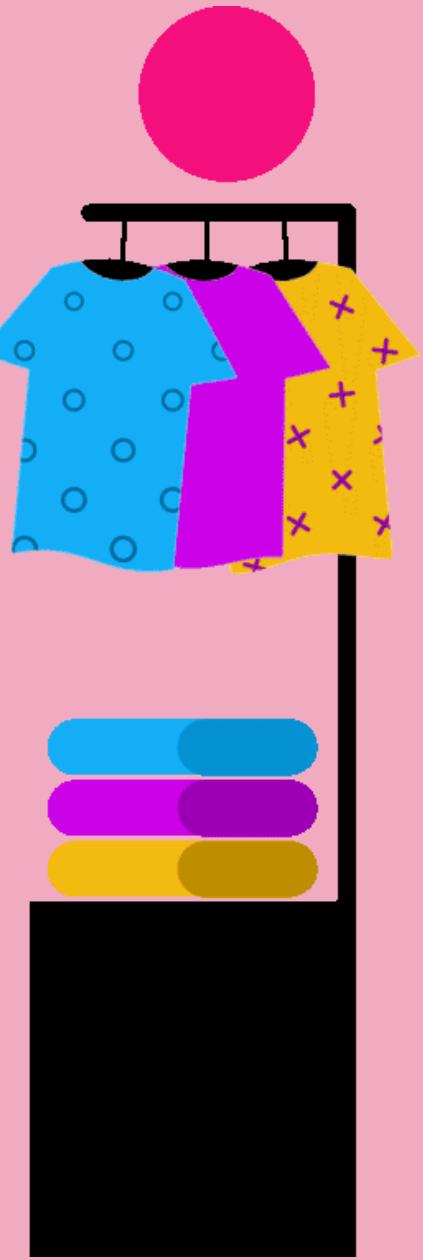
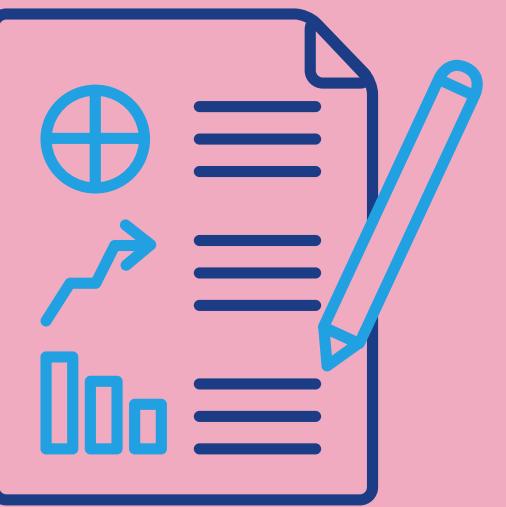
Definition

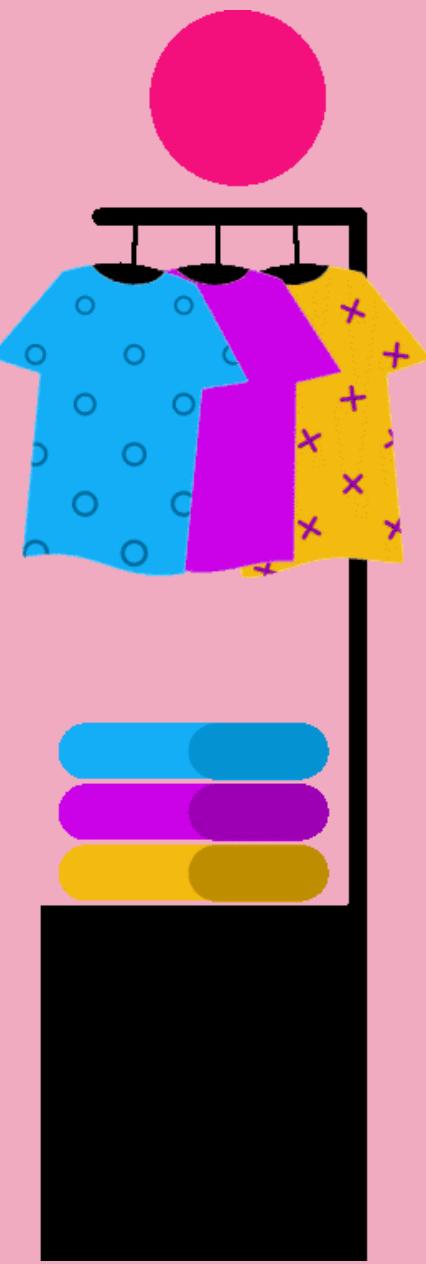
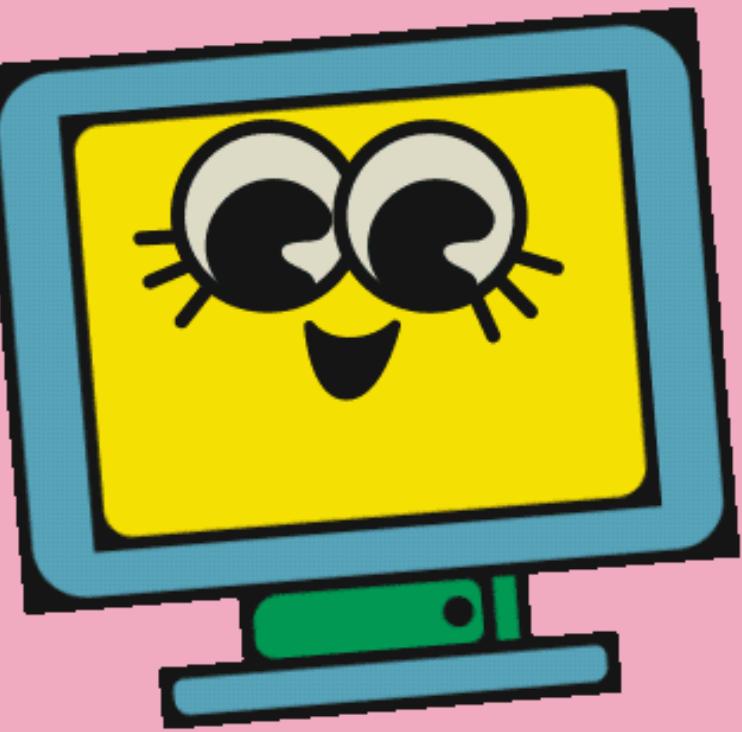
תחום מדעי הנתונים
כולל בתוכו שיטות
מדעית, אלגוריתמים
וממערכות שматרתם
להפיק ידע מנ נתונים
גולמיים.







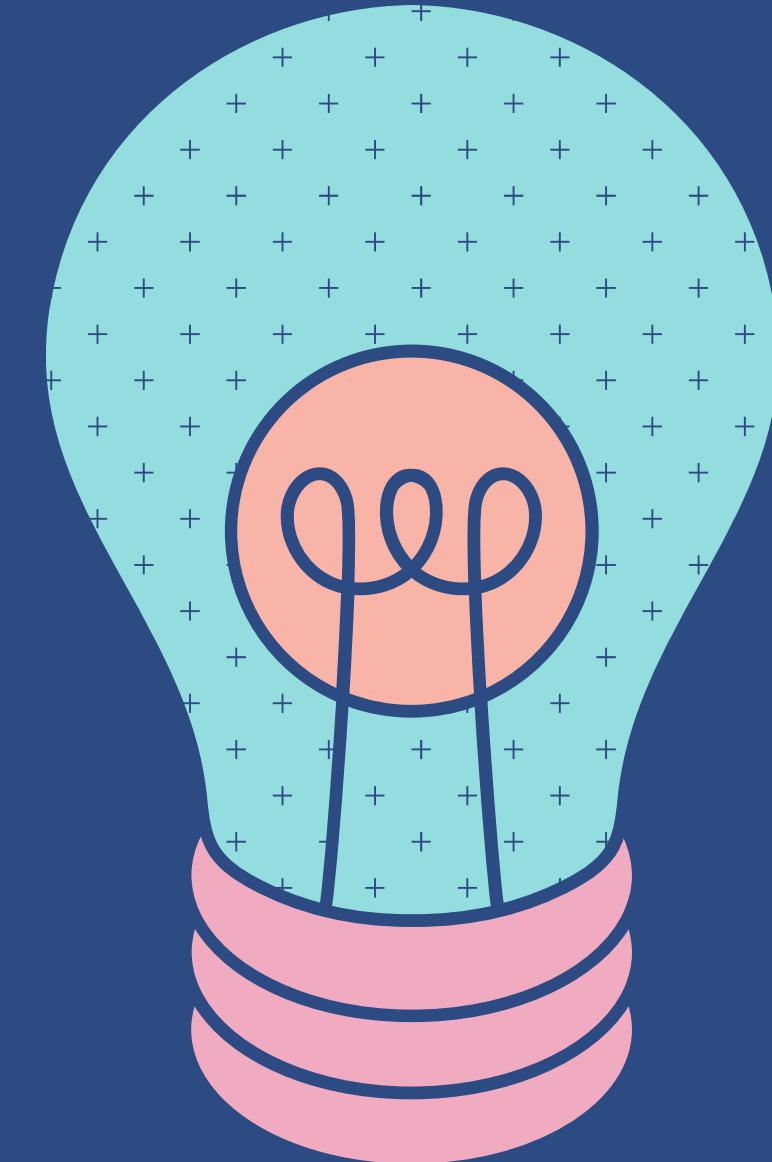




What is AI?

”לגרום למכונה להתנהג בדרך
שהייתה נחשבת לאינטיליגנטית לו
אדם התנהג כך.”

מרווין לי מינסקי (1927 - 2016) זוכה פרס טיוריינג



מה זה מודל למידת מכונה?

מודל הוא קובץ קוד שאומן באמצעות סט
נתוניים כדי לזהות תכניות בתוך הנתוניים.

ניתן להשתמש במודל שכבר אומן על מנת לקבל חיזוי על נתונים
שהוא טרם ראה.



A person's arm and hand are visible in the foreground, reaching towards a computer monitor. The monitor displays a terminal window with code snippets in Python and JavaScript. The code appears to be related to OAuth authentication for Twitter, specifically for getting user access tokens and tweets. The background is dark, and the monitor is the primary light source.

```
t_tweets_payload.py JS get-user.js x JS get-tweet.js
ts:cURL > JS get-user.js > accessToken
new URL('https://api.twitter.com/oauth/access_token');
new URL('https://api.twitter.com/oauth/authorize');
new URL('https://api.twitter.com/labs/1/users');

becker',
    prompt) {
      sync (resolve, reject) => {
        prompt, (out) => {
          ;
        }
      }
    }

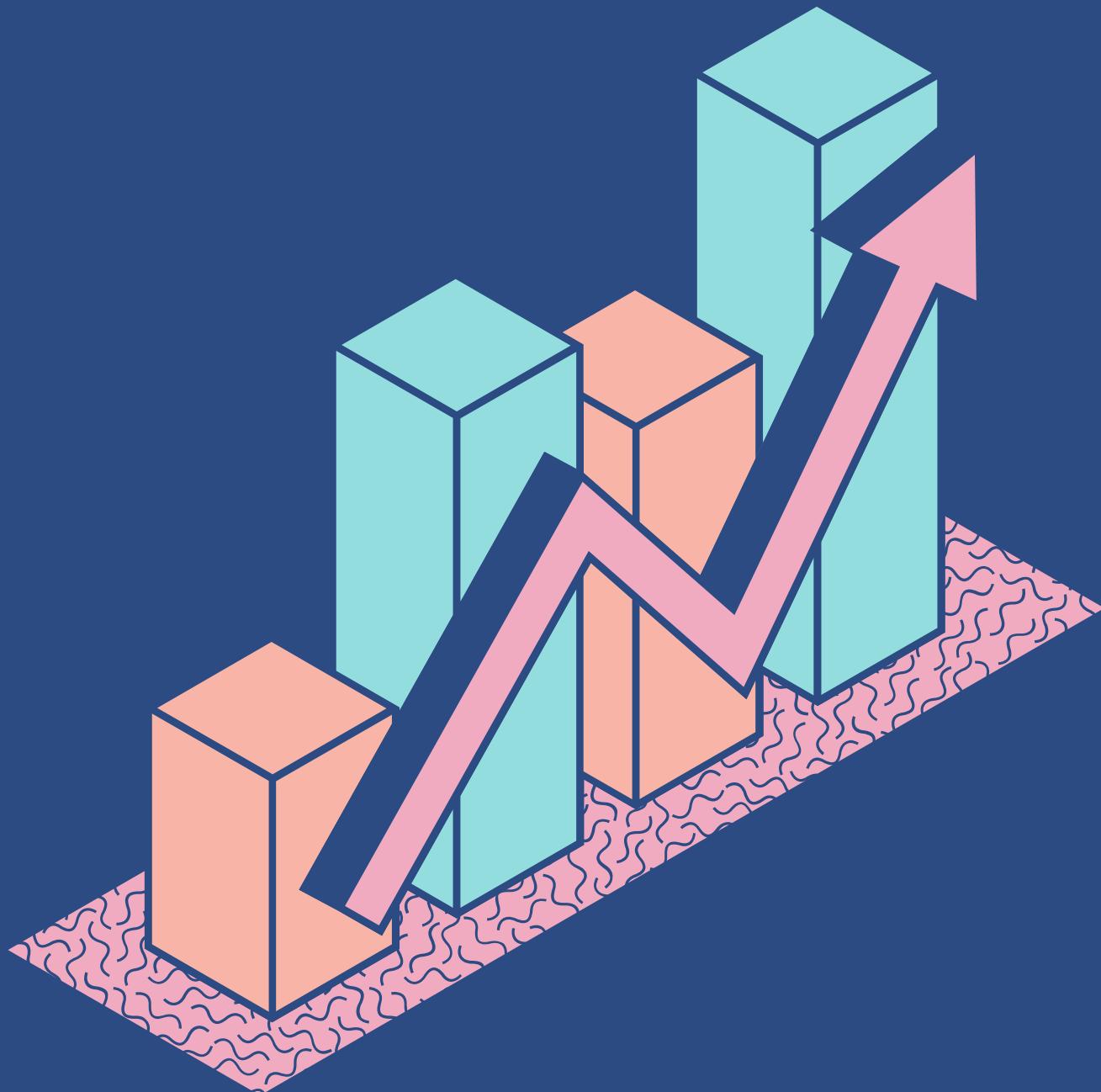
oken({oauth_token, oauth_token_secret}, verifier) {
  {
    consumerKey,
    consumerSecret,
    token,
    token_secret,
    verifier,
    verifierToken,
    verifierTokenSecret
  }

st({url: accessTokenURL, oauth: oAuthConfig});
q.body);
annot get an OAuth request token');

Token() {
  {
    consumerKey,
    consumerSecret,
    verifier,
    verifierToken,
    verifierTokenSecret
  }

post({url: requestTokenURL, oauth: oAuthConfig});
q.body);
```

דוגמה למודל Random Forest





עץ החלטה - מקבל החלטה עברו הנתונים שקיבל





עיר רנדומי - מקבל את החלטת הרוב

שלבי תحلיל ה-Data Science



מוכנות לדוגמה?



- טיטanic היה אוניית ענק שטבעה בהפלגת הבכורה שלו ב-1912
- שאלת המחקר שלנו: מי ישרוד את שקייטה של הטיטanic?
- הנתונים שלנו מכילים טבלה עם כ-198 רשומות ובכל רשומה מאפיינים על אחד מנוסעי הטיטanic
- נבע עיבוד על הנתונים, כך שהמודל יוכל להתאים עליהם ולחזות את משתנה המטרה: Survived - בן או לא בן
- נעריך את המודל ואת אחוז הדיווק שלו - האם הוא צדק או טעה?





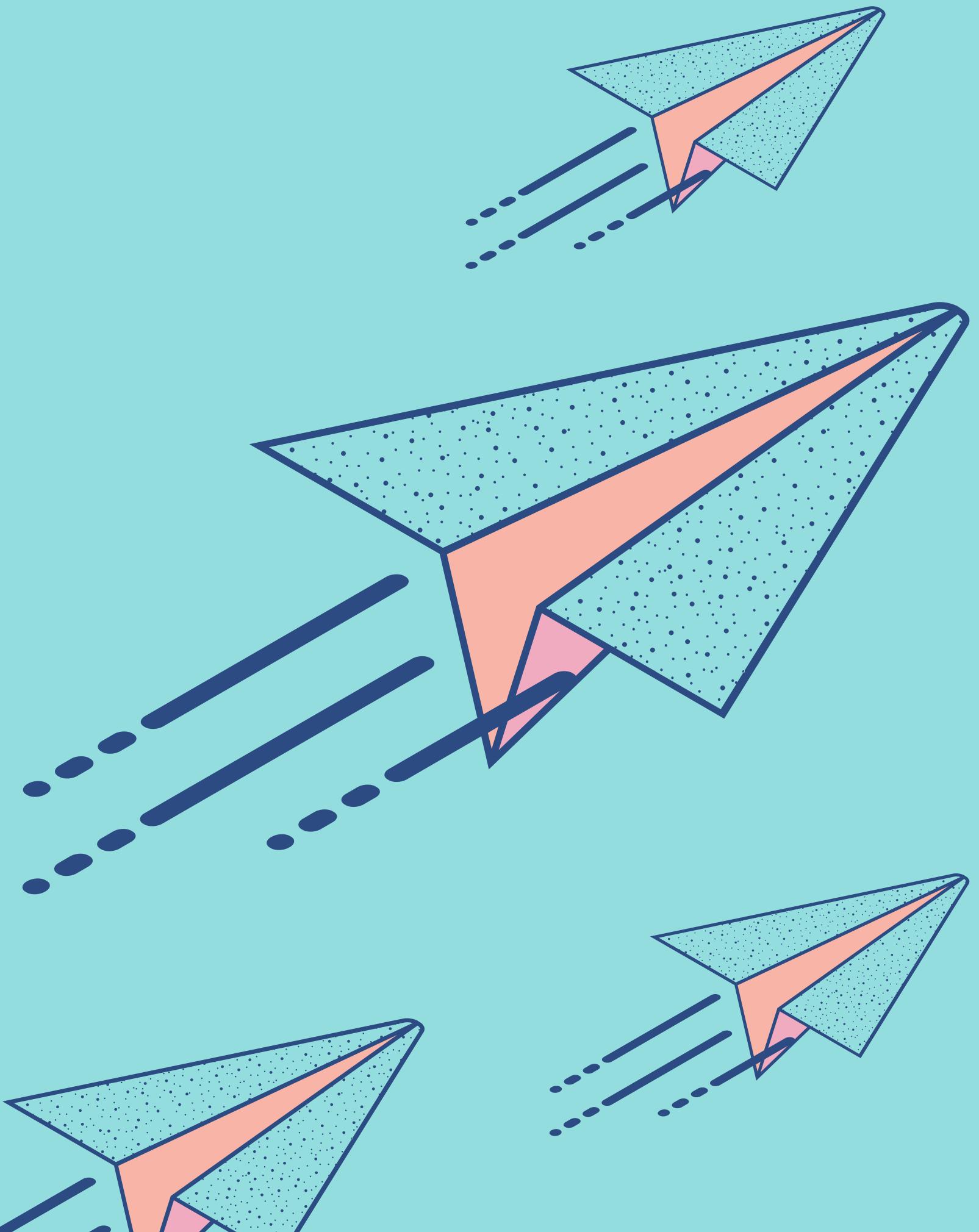
הנחיות:

- היכנסו לינק: [she_codes_data_science_session](#)
- לחצו על הלחצן הירוק 'code' ולאחר מכן 'download zip'
- פתחו את ה-zip במחשב ושמרו את שני הקבצים בתוך ה-`google drive` שלכם (בעמוד הראשי)

מי שמכנה יכולה לקחת כמה דקות הפסקה ומיהיד נתחיל



שאלות?



סיכום

או מה היה לנו היום?

- **למדנו על תחום מדעי הנזוטים**
- **עברנו ביחד על שלבי תהליכי הפקת הידע מנתוניים גולמיים**
- **למדנו על בינה מלאכותית ועל מודל למידת מכונה Random Forest**
- **ראינו כיצד מודל מצטlich לענות על שאלת מי ישרוד את שקיעת הטיטא尼克 בדיקן של מעלה מ-70%!**

