**שאלות החובה בנושא בינה מלאכותית :**

**1. איך למידת המכונה בפרויקט משתלבת בשאר הפרויקט, באיזה חלקים של הפרויקט אתם צריכים מכונה לומדת?**

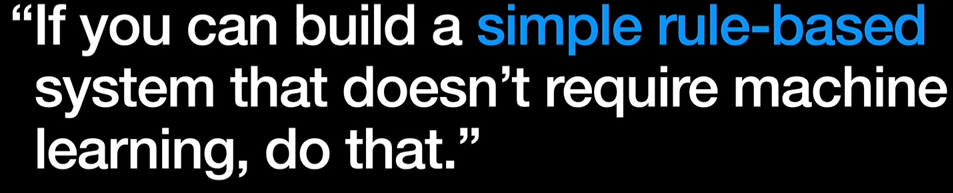
למידת המכונה בפרויקט שלנו משתלבת בפרויקט שלנו בשני חלקים מרכזיים: זיהוי אות , ולאחר מכן סיווגה של האות. בינה מלאכותית אחת תצטרך לזהות בתוך התמונה את האות , ובינה מלאכותית שנייה תצטרך לסווג את האות בתוך התמונה, כלומר לומר מהי האות שצולמה.

**2. למה אתם צריכים למידת מכונה בפרויקט, הסבירו למה אתם לא עושים זאת בתכנות רגיל?**

אנו צריכים למידת מכונה בפרויקט , משום שביצוע זאת בעזרת עיבוד תמונה רגיל הוא מאוד מורכב ודורש המון מידע ולא בטוח שאפשרי, זו תהיה משימה כמעט בלתי אפשרית לכתוב קוד בתכנות רגיל. ביצוע הפרויקט בבינה מלאכותית יהיה יותר יעיל משום שאנו משתמשים במאגר מידע של אותיות מוכנות ומלמדים את המכונה לזהות אות ולסווג אותה.

3. **מה החוק מספר 1 של למידת מכונה?**

"אם אתה יכול לבנות מערכת בעלת חוקים בסיסיים שלא צורכת למידת מכונה , תעשה את זה , זה עדיף! "



https://media.discordapp.net/attachments/659029109258584085/739872700780249218/unknown.png

**4. לאיזה קטגוריה של למידה המכונה הלומדת שלכם מתאימה (סוג הלמידה, עדיפות לפירוט גם על המשפחה של האלגוריתמים המתאימים)?**

הקטגוריה היא למידה מפוקחת- Supervised Learning:

כאשר יש לנו מאגר של נתונים ותשובות שהושגו בצורה כזו או אחרת (צילומים, מדידה, עבודה של מומחים, וכו’) – ואנו רוצים שהמחשב ינסה לחקות את התשובות.

במקרה שלנו, יש לנו מאגר מידע שמכיל בתוכו תצלומים של אותיות ומבין שעליו לזהות שמדובר באות ומכאן גם לסווג אותה בהמשך.

**5. עם איזה מאגרי מידע תעבדו, אם אין לכם מאגר, מול מה ואיך המכונה תלמד?**

אנחנו נעבוד עם מאגר מידע של תווים ואותיות שמתאים לסוג פרויקט של OCR, מאגר של OCR.

**6. איך תכינו את המידע לצורך שלב העיבוד של המכונה הלומדת?**

באמצעות מאגרי המידע של אותיות ותווים המותאמים ל- OCR.

**7. לפי איזה קריטריונים המכונה הולכת ללמוד, מה המטרה של פונקציה מפעילה?**

במקרה של זיהוי אוביקטיים/מצב בתמונה בדרך כלל נרצה לצמצם את הבעיה כמה שיותר, ולכן נעשה קודם כל תהליך של צמצום הבעיה לפי פרמטרים מסוימים שקשורים למה שנרצה לזהות, לדוגמא סינון של כל הפיקסלים שבטווח צבעים מסוים ואת כל השאר לא להכניס.

הקריטריון הבא קשור לצורת המימוש, נשתמש במודל CNN כי בדרך כלל זה מה שהולכים אליו ישר כאשר יש בעייה הכוללת תמונה.

**8. איך תבדקו שהתשובות של המכונה הלומדת "טובות"?**

נבחן את הבינה המלאכותית בכך שנציג בפניה תמונה והיא תצטרך לזהות את האותיות ולאחר מכן לסווג אותם , מכאן נראה אם המכונה הצליחה לזהות ולסווג בצורה המיטבית שלה את האות ומכאן נידע שהתשובות של המכונה הלומדת "טובות". בעיקרון , השוואה של התשובה של המודל לתשובה הידועה מראש.

**איך יתאפשר אימון מחדש של המכונה מידי פעם בלי לקמפל מחדש את כל הפרויקט?**

את הנתונים נטען באמצעות פונקציית save שמספקת ספריית TensorFlow.

**10. באיזה כלי פיתוח אתם הולכים להשתמש?**

מודל CNN .

TensorFlow, שהוא ספרייה בשפת פייתון ובכך אימון המודל נעשה על סט נתונים מפורסם ששמו emnist הכולל תמונות של תווים.

**11. הסבר במילים לפי מה המכונה לומדת בדיוק ומה גורם לה לעבוד?**

המכונה לומדת לפי התמונות ומאגר המידע שלה ומה שגורם לה לעבוד הוא התמונה שהלקוח מספק לה.

**12. באיזה סביבת עבודה אתם הולכים לעבוד?**

סביבת העבודה שלנו הוא פייתון ונשתמש בספרייה ששמה TensorFlow

**13. איזה תחומי מתמטיקה רלוונטיים לפרויקט שלכם?**

רגרסיה קווית (לינארית) משמשת למציאת הקשר בין נתונים מספריים.

**14. איך ואיפה ובמי תעזרו כדי ללמוד את המתמטיקה הרלוונטית?**

באמצעות ספריית TensorFlow 2. <https://reshetech.co.il/machine-learning-tutorials/intro-to-tensorflow-2-or-how-to-find-linear-regression>

**15. באיזה מקורות מידע תעזרו בפרויקט (לא שם של אתר, אלא לינק למאגר עצמו שרלוונטי במדויק לתחום הפרויקט שלכם)?**

**1)**

<https://reshetech.co.il/machine-learning-in-the-browser/how-to-develop-hand-written-digits-recognition-web-app>

**2)**

<https://towardsdatascience.com/a-gentle-introduction-to-ocr-ee1469a201aa>

**16. איך ואיפה אתם הולכים ללמוד תכנות של מכונה לומדת, (גם כאן, לינק כמה שיותר ספציפי)?**

<https://reshetech.co.il/machine-learning-tutorials/all-the-tutorials>

**17. האם יש לכם מחשב המסוגל לבצע אימון של מכונות לומדות? אם לא, באיזה שירות ענן או משאבים חיצוניים תשתמשו?**

יש לנו מחשב המסוגל לבצע אימון של מכונות לומדות, במידה ולא יעבוד נשתמש אפשרות נוספת היא להשתמש בסביבת colab, שמאפשרת לבצע למידת מכונה על הענן של גוגל בחינם, כל מה שצריך הוא חשבון GMAIL פעיל, ניתן לעבוד על git hub אותו למדנו.

**18. במידה ותשתמש בשירות ענן, איך תלמדו איך לעבוד מולו? (גם כאן, לינק ספציפי ככל האפשר)**

<https://www.tensorflow.org/guide/distributed_training?hl=he>

**19. תתנו לנו לינקים למאמרים שבהם תעזרו, הסבירו לכל מאמר מדוע הוא קשור לפרויקט שלכם?**

**1)**<https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5>

<https://www.youtube.com/watch?v=aircAruvnKk&feature=emb_logo>

<https://www.onlineocr.net/>

**20. במידה ותגלו שהדאטה סט שלכם לא טוב, האם יש לכם לינקים לדאטה סטים אחרים שאולי רלוונטים?**

[**https://lionbridge.ai/datasets/15-best-ocr-handwriting-datasets/**](https://lionbridge.ai/datasets/15-best-ocr-handwriting-datasets/)**1)**

<https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research#Reviews> **2)**

**3** <https://www.nist.gov/itl/products-and-services/emnist-dataset> **(**

**21. איך תדע מה רלוונטי באמת ללמוד מבין הדברים שאתה מתכנן ללמוד?**

לפי האינטרנט, נצטרך ללמוד לזהות את האובייקט והאות בתמונה באמצעות הבינה מלאכותית ומשם לסווג אותו.