**חיזוי מזג אוויר ליום הבא בעיר ניו יורק ע"פ נתונים היסטריים**

משתתפים

ברק זן, 305634487, [barakzan@campus.technion.ac.il](mailto:barakzan@campus.technion.ac.il)

אופיר פיצי שומרון, 201450574, [ofirshomron@campus.technion.ac.il](mailto:ofirshomron@campus.technion.ac.il)

תיאור הבעיה

ישנן כמה בעיות שברצוננו לנסות לפתור, נציג את כולן ובמהלך התקדמות הפרוייקט ונחליט בעזרת המנחה לאיזה בעיות כדאי לגשת.

הבעיות שמעניינות אותנו הן ולא בהכרח רק:

* האם מחר יהיה יותר חם או יותר קר ביחס להיום
* האם ירד מחר גשם או לא
* חיזור רמת העננות מחר: בהיר\מעונן חלקית\מעונן\קודר
* האם תהייה מחר סופה או לא

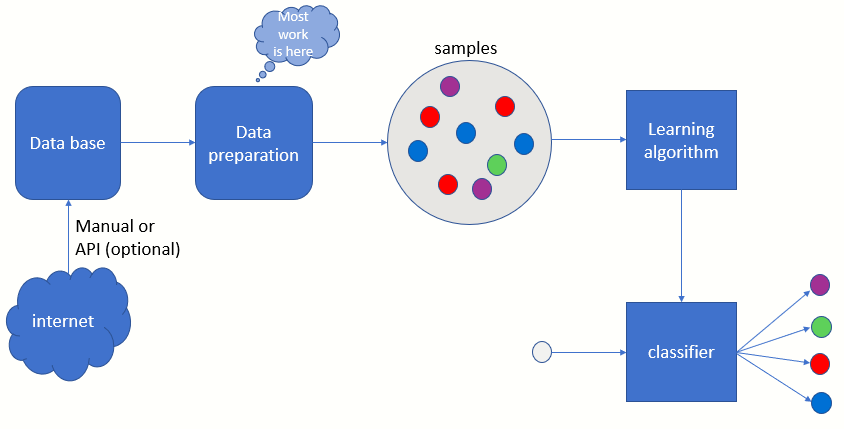
הצעת פתרון

הפתרון יתחלק לשלושה שלבים עיקריים:

* השגת היסטוריית מזג האוויר של ניו יורק מהאינטרנט
* ביצוע אגריגציה למידע כך שיתאים לשימוש באלגוריטמי למידה.
* שימוש בגרסא מורחבת של KNN ע"מ לבצע חיזוי.
* ביצוע ניסויים אוטומטיים ע"מ לשפר את דיוק האלגוריתם ע"י בדיקת פרמטרים.

המידע בו נשתמש יהיה היסטורית מזג האוויר בעיר ניו יורק, בנוסף קיימת אופציה להסתכלות על מזג אוויר בערים שכנות בשילוב עם הסתכלות על כיוון הרוח ע"מ לקבל חיזוי מדוייק יותר.

תיאור המערכת



\*קיימת אפשרות שנעשה אוטומציה שמקבלת עיר ומורידה את המידע הרלוונטי אליה.

החלק הראשון **והמשמעותי ביותר** של המערכת, הוא data preparation, בחלק זה ניקח את המידע שנשיג מהאינטרנט וניצור ממנו סט דוגמאות למידה ע"י:

* מילוי מידע חסר
* ביצוע אגריגציה
* ניקוי נעשים
* נרמול

החלק השני יהיה חלוקת הדוגמאות שקיבלנו ל3 קבוצות:

* קבוצות מבחן – כל הימים מתחילת 2018
* קבוצת הערכה – כל הימים בשנת 2017
* בקוצת אימון – הימים שקדמו ל2017

\*הגדלים הנ"ל אינם סופיים או מחייבים.

החלק השלישי יהיה שימוש באלגוריתם KNN הממוש כבר בסיפריות קיימות וקבלת מסווג לחיזוי.

ע"מ לשפר את דיוק האלגוריתם נבצע כיוון פרמטרים ע"י כתיבת אוטומציה שתבדוק את תוצאות אלגוריתם הלמידה לאחר מתן משקלים שונים לכל תכונה ובדיקת התוצאות לכל סט משקלים נתון על קבוצת ההערכה.

הערכת ביצועי המערכת

בכדי להעריך את ביצועי המערכת, נסתכל על מטריצת מטריצת הבלבול ונחשב את אחוז השגיעה ואחוז הדיוק. חשוב לציין, שבמקרה שלנו, אין סוג שגיאה ספציפי החמור יותר מאחר (לדוגמא להגיד שחולה בריא)

פרטים טכניים

שפת תכנות: פייתון 3

סביבת עבודה: pycharm

ספריות ומודולים: sklearn, pandas, numpy, scipy ועוד שנוסיף בהתאם להתקדמות הפרוייקט