הצעת פרויקט למידת מכונה

שם המשתתפים: דביר סעדון

תיאור של המאגר: המאגר מתאר מידע לגבי יין אדום מסוג "Vinho Verde"

המאגר מכיל 12 פיצ'רים: הם מכילים בינהם כאלה שמתארים כמות של סוגי חומצה שונים שנמצאים ביין כמו גם כמו גם סוכר ומלח. בנוסף, כמות הגופרית הדו חמצנית ביין, ריכוז האלכוהול, צפיפות היין ו-PH והמשתנה שארצה לחזות – האיכות של היין מ0-10.

השאלות שעליהם ארצה לענות: ארצה לענות על השאלה של חיזוי (classification) של הדירוג איכות של היין (0-10). אנסה גם למצוא את המשוואה לregression של האיכות. אנסה לענות על אילו הם הפיצרים הכי טובים כדי לענות על השאלה הראשונה.

הטכניקות בהם אשתמש: Random Forest, KNN, Adaboost וגם Decision tree.

בכדי לענות על השאלה השנייה, אשתמש בLogistic regression.