

הצעת פרויקט בלמידת מכונה

מגישים:

יונתן בוריצקי (207254194)

טל טוסיאנו (208846600)

תיאור המאגר:

המאגר שלנו מורכב מכמות גדולה של צילומי מסך מתוך כמה שיותר עמודי אינסטגרם אשר ירחיבו את המאגר ואת כמות הטסטים לסיווג שנוכל לבצע. כך בעצם נגדיל את הסיכוי לסיווג טוב ולתוצאה המירבית

מקור המאגר:

המאגר לקוח מתוך אתר Kaggle והנתונים בו נאספו באופן ידני מתוך חשבונות Instagram

קישור למאגר: [Instagram Page Screen Shots in 5 Category](#)

השאלות שעליהן נרצה לענות:

אנו נרצה לסווג את צילומי המסך ל-5 קטגוריות ולכן השאלות שעליהן נרצה לענות הן-

- 1) האם עמוד אינסטגרם זה עוסק ביופי? כלומר בעל העמוד הוא בלוגר בענייני יופי
- 2) האם עמוד אינסטגרם זה עוסק באופנה? כלומר בעל העמוד הוא בלוגר אופנה
- 3) האם עמוד אינסטגרם זה עוסק בכושר? כלומר בעל העמוד הוא מדריך כושר/בלוגר לאיכות חיים
- 4) האם עמוד אינסטגרם זה עוסק באוכל? כלומר בעל העמוד הוא בלוגר של אוכל
- 5) האם עמוד אינסטגרם זה עוסק במשפחה? כלומר בעל העמוד הוא איש משפחה

הטכניקות שנשתמש בהן:

Clustering היא שיטה המשמשת לקיבוץ אובייקטים דומים בהתבסס על המאפיינים או התכונות שלהם. אנו נוכל להשתמש לצורך סיווג ודחיסת תמונות. נקבץ תמונות דומות על סמך התכונות החזותיות שלהן כגון צבע, מרקם וצורה. בנוסף לקבץ פיקסלים דומים ולייצג אותם עם ערך בודד.

CNN הם סוג של מודל למידה עמוקה שיכול לשמש למשימות סיווג תמונות. הם פועלים על ידי לימוד קבוצה של מסננים שיכולים לחלץ תכונות מתמונות, ולאחר מכן שימוש בתכונות אלה כדי ליצור תחזיות לגבי קטגוריית התמונה.

SVM מודל זה מבוסס על הפרדה של וקטורים עם מרווח מקסימלי. מודל זה מאפשר הפרדה של דאטא מממד גבוה ולכן יכול להתמודד עם תמונות בעלות אלפי פיקסלים.

Random Forests יערות אקראיים הם סוג של אלגוריתם למידת אנסמבל שניתן להשתמש בו למשימות סיווג. הם עובדים על ידי בניית עצי החלטה מרובים ושילוב התחזיות שלהם כדי ליצור תחזית סופית.

AdaBoost עובד על ידי שילוב של מסווגים "חלשים" מרובים למסווג "חזק" אחד. כל מסווג חלש מאומן על תת-קבוצה של הנתונים, והאלגוריתם מקצה משקלים גבוהים יותר לנקודות הנתונים שסווגו בצורה שגויה. לאחר מכן, המסווג החלש הבא מאומן על מערך הנתונים המעודכן, כאשר המשקולות משקפות את הקושי בסיווג כל נקודת נתונים. תהליך זה חוזר על עצמו במשך מספר קבוע של איטרציות או עד ששיעור השגיאות ממוצע.

לאחר שכל המסווגים החלשים אומנו, הם משולבים למסווג אחד חזק באמצעות הצבעת רוב משוקלל. המשקלים של כל מסווג חלש נקבעים על סמך דיוקם, כאשר למסווגים מדויקים יותר ניתן משקל גבוה יותר.

