



מאיקס עיגול ושחמט ועד תרגום שפות – למידת מכונה

למידת מכונה הופכת יותר ויותר לכלי חשוב בעולמות ההשקעות, הבנקאות, הביטוח והפנסיה – עד כדי כך שכמעט כל העובדים צפויים להיות מושפעים ממנה בצורה כזו או אחרת במהלך השנים הבאות. למידת מכונה עושה שימוש במאגרי נתונים גדולים Big Data על מנת ללמוד על קשרים בין משתנים, לבצע תחזיות ולפעול באינטרקציה עם סביבה משתנה. מאגרי נתונים גדולים על משתנים המתארים רכישות של צרכנים, תנועות של מחירי מניות והרבה היבטים אחרים של עסקים אינם חדשים כיום. השיפורים הטכנולוגיים האחרונים כגון העליה במהירויות המחשב והירידה בעלויות אחסון הנתונים מאפשרים לנו להגיע למסקנות מתוך מאגרי נתונים גדולים בדרכים שפשוט לא היו אפשריים לפני 20 או 30 שנה.

למידת מכונה היא ענף של בינה מלאכותית (AI). בינה מלאכותית מפתחת דרכים שבהם למידת מכונה יכולה לחקות בינה אנושית, ואפילו לחקות אותה. למידת מכונה יוצרת בינה על ידי למידה מתוך נפחים גדולים של נתונים. למידת מכונה היא ללא ספק ההתפתחות המרגשת ביותר בתוך הבינה המלאכותית והיא זו שיש לה את הפוטנציאל הרב ביותר לשנות עסקים.

על מנת להסביר כיצד למידת מכונה נבדלת מגישות אחרות של בינה מלאכותית ניקח את המשימה הפשוטה ביותר של לתכנת מחשב לשחק איקס עיגול. גישה אחת היא לספק למחשב טבלת בדיקות המציגה את הפוזיציות שיכולות לקרות ואת המהלך שהיה מבוצע על ידי שחקן אנושי מומחה עבור כל אחת מאותן פוזיציות. גישה אחרת היא להציג למחשב מספר גדול מאוד של משחקים (למשל, על ידי כך שניתן למחשב לשחק מול עצמו אלפי פעמים) וניתן לו ללמוד את המהלך הטוב ביותר. הגישה השנייה הינה יישום של למידת מכונה. למעשה ניתן להשתמש בהצלחה בכל אחת מהגישות שתוארו לצורך משחק פשוט כמו איקס עיגול. גישות



של למידת מכונה הוכחו ככאלה שעובדות מצויין עבור משחקים יותר מורכבים כמו שחמט וגו (למי שלא מכיר מדובר במשחק אסטרטגיה מופשט שמקורו בסוף) כאשר הגישה הראשונה בבירור אינה אפשרית.

אחת הדוגמאות הטובות לעוצמה של למידת מכונה היא תרגום שפה. כיצד ניתן לתכנת מחשב לתרגם בין שתי שפות, נגיד מעברית לגרמנית? רעיון אחד שעולה לי לראש הוא לתת למחשב מילון עברי-גרמני. למרבה הצער תרגום מילה-למילה מספק תוצאות גרועות מאוד ולכן הכרחי לנסות ולתכנת גם את כללי הדקדוק העברי והדקדוק הגרמני. זוהי מלאכה ממש לא פשוטה וגם אחרי שהיא נעשית הרי שתוצאותיה עדיין רחוקות מלהיות מושלמות. גוגל היא חלוצה בשימוש באלגוריתמים של למידת מכונה לצורך תרגום. גוגל הכריזה על כך בנובמבר 2016 והיא קוראת לטכנולוגיה החדשה בשם GNMT (תרגום המכונה העצבי של גוגל, Google Neural Machine Translation). במסגרת טכנולוגיית ה- GNMT נותנים למחשב מילון דפים של חומר שתורגם על ידי מומחי תרגום מעברית לגרמנית. המחשב לומד את החומר ומפתח כללי תרגום משלו. התוצאות מטכנולוגיית ה- GNMT מהוות שיפור גדול בהשוואה לגישות הקודמות.

מדע נתונים (Data Science) הוא תחום שכולל בחובו למידת מכונה ולעיתים נחשב לרחב יותר וכולל גם משימות כמו למשל קביעת מטרות, יישום והטמעת מערכות ותקשורת עם בעלי עניין. נציין כי יש שרואים במונחים "למידת מכונה" ו- "מדע נתונים" כמילים נרדפות, מאחר וקשה לראות כיצד מומחי למידת מכונה יכולים להיות יעילים לארגון כלשהו אם רק עושים שימוש בדאטה (Big Data) באמצעות בניית מודלים, פיתוח ושימוש באלגוריתמים וניתוח תהליכים לצורך זיהוי כיוונים ומגמות במגוון תחומים לרוחב הארגון מבלי: (1) להיות אחראים על זיהוי אתגרים עסקיים בהם DATA יכול להוות גורם מכריע בשיפור קבלת החלטות; (2)



להיות אחראים על איתור ואיסוף מקורות מידע פנים ארגוניים וחיצוניים, הגדרה ואיפיון של שימושי המידע בארגון; (3) לבנות מאגר המידע שאינם מובנים ונמצאים במסגרת הארגון, אפיון והגדרת הצגת המידע ותוצריו לדרג מקבלי ההחלטות בארגון; (4) לפתח כלים, מודלים, תהליכים ומערכות לרוחב הארגון בתחום האנליזה; ו- (5) לנהל מעקב אחר התקדמות הפרויקטים בהתאם לתכנית העבודה אל מול מנהלי התחומים והעובדים. הוצה לומר- אף ארגון לא באמת יעסיק מומחי למידת מוכנה בשביל שאלו רק יעבדו עם מאגרי נתונים גדולים תוך פיתוח ושימוש באלגוריתמים של למידת מכונה על מנת לזהות כיוונים ומגמות בעולמות תוכן שונים.

ניתן לתאר את למידת המכונה או מדע הנתונים כעולם החדש של סטטיסטיקה. הסטטיסטיקה המסורתית נוגעת לנושאים כמו הסתברויות, התפלגויות, רווחי סמך, מבחני מובהקות ורגרסיה לינארית. נאמר מראש ידע בנושאים הללו הוא חשוב, אבל כיום אנו יכולים ללמוד מתוך מאגרי נתונים גדולים בדרכים שלא היו אפשריות לפני כן. לדוגמא, אנו יכולים לפתח מודלים לא-לינאריים לניבוי ושיפור קבלת החלטות, לחילופין אנו יכולים לחפש דפוסים בנתונים על מנת לשפר את ההבנה של החברה את לקוחותיה והסביבה שבה היא פועלת או לחילופין חילופין אנו יכולים לפתח כללי החלטה כאשר אנו פועלים באינטרקציה עם סביבה משתנה.

כאמור, יישומים אלו של למידת מכונה אפשריים היום הודות לשיפורים הטכנולוגיים האחרונים כגון העליה במהירויות המחשב והירידה בעלויות אחסון הנתונים.

ונחתום בעדות אישית. כסטטיסטיקאי וכאקונומטריקאי בעת טבילת האש הראשונה שלי בתחום למידת המכונה נחשפתי לטרמינולוגיה שנראתה לי משונה. לדוגמא, אני כסטטיסטיקאי וכאקונומטריקאי מדבר על משתנים בלתי תלויים ומשתנים תלויים בעוד

שמדעני נתונים מדברים על מאפיינים (Features) ויעדים (Targets). למעשה למדתי שפה חדשהואני לא מתכוון ל- Python ולספריות הסטטיסטיות שלה כמו NumPy (שעובדת עם מערכים וקוראת קבצי טקסט), Matplotlib (שיוצרת תרשימים), Scipy (שמבצעת אופטימיזציה ופותרת משוואות א-לינאריות) ו-Pandas (תיקיית האלגוריתמים של למידת מכונה).

פרטים אודות כותב המאמר: האקטואר רועי פולניצר, FRM

רועי בעל תואר שני במימון (התמחות בניהול סיכונים ואקטואריה) ותואר ראשון בכלכלה (התמחות במימון), שניהם מאוניברסיטת בן-גוריון בנגב, בעל דיפלומה בניהול סיכונים פיננסיים (FRM®) מאוניברסיטת אריאל בשומרון ולמד בתוכנית ללימודי תעודה באקטואריה באוניברסיטת חיפה. כמו כן, רועי אקטואר מלא



(Fellow) בלשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (F.I.L.A.V.F.A.), מוסמך כמעריך שווי מימון תאגידי (CFV) מטעם לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (IAVFA), מוסמך כמנהל סיכונים פיננסיים (FRM) מטעם האיגוד העולמי למומחי סיכונים (GARP) ומוסמך כמומחה לניהול סיכונים (CRM) מטעם האיגוד הישראלי למנהלי סיכונים (IARM).

לרועי ניסיון של מעל ל- 15 שנה בביצוע ניתוחים כמותיים במכשירים פיננסיים, בהערכת שווי תאגידים ונכסים בלתי מוחשיים, באמידה וכימות סיכונים כמו תמותה, אריכות ימים, תחלואה, ביטולים והחלמה מנכות, ובמידול ומדידת סיכונים שוק, אשראי, תפעוליים, מודל, נזילות והשקעות לצורכי יישום הוראות רגולטוריות ותקינה חשבונאית, פיתוח, יישום ותיקוף מדלים בתחומים של הערכות שווי, ניהול סיכונים, אקטואריה והנדסה פיננסית, קביעת תעריפי ביטוח חיים, הערכת פרמיות סיכון והערכת עתודות ביטוח, קביעת עלות תנאי פנסיות (צוברות



ותקציביות) והכנת מאזנים אקטואריים לקרנות פנסיה, ניתוח וחיזוי מצבים פיננסיים מורכבים וכן העברת סמינרי הדרכה והשתלמויות בתחומי התמחותו: מימון, אקטואריה, הערכות שווי, בנקאות, ניהול סיכונים, אופציות והנדסה פיננסית.

ניסיונו של רועי בתחום ה-Data Analysis, כולל: עבודה עם מאגרי מידע גדולים Big Data תוך שימוש ב-Statistical Learning (כגון: סטטיסטיקה תיאורית, הסתברות, הסקה סטטיסטית, סטטיסטיקה א-פרמטרית, חלוקת נתונים, נרמול נתונים, Fitting ו-Bayes Theorem) ובאלגוריתמים מסוג Unsupervised Learning (כגון: Hierarchical, k-means Clustering, Density-based Clustering, Clustering, Distribution-based Clustering ו-Principle Components Analysis) למציאת דפוסים וזיהוי מגמות ואנומליות בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, פיתוח תשתית לצורך ניתוח נתונים, שילוב והטמעת כלים לצורך גישה ושליפה עצמאית של נתונים ממאגרי מידע, פיתוח דוחות, ממשקים ומסכים באמצעות כלי ויזואליזציה.

ניסיונו של רועי בתחום ה-Data Science, כולל: עבודה עם מסדי נתונים גדולים Big Data תוך שימוש באלגוריתמים מסוג Supervised Learning (כגון: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Logistic Regression, Maximum Likelihood Estimation, k-Nearest Neighbors, Decision Tree, Random Forest, Ensemble, Bagging, Boosting, Naïve Bayes Classifier, Linear Separation, Support Vector Machine, Non-Linear Separation, SVM Regression, Artificial Neural Network, Convolutional Neural Network ו-Recurrent Neural Network) לניבוי וסיווג בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה ובמודלים מסוג Reinforcement Learning (כגון: Q-learning, Monte Carlo, Simulation, Temporal Difference Learning ו-n-Step Bootstrapping) לקבלת החלטות מרובות שלבים בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, זיהוי אתגרים



עסקיים שבהם DATA יכול להוות גורם מכריע בשיפור קבלת החלטות, איתור ואיסוף מקורות מידע, הגדרה ואיפיון של שימושי המידע, בניית מסד המידע, אפיון והגדרת הצגת המידע ותוצריו, פיתוח כלים, מודלים, תהליכים ומערכות בתחום האנליזה, תוך שימוש בכלי אנליזה מתקדמים (EXCEL, VBA ו-R).