## מבוא ללמידת מכונה (Machine Learning)

למידת מכונה היא ענף של בינה מלאכותית העוסקת בלמידה מתוך מאגרי מידע גדולים ( Big Data). למידת מכונה כרוכה בפיתוח וגיבוש אלגוריתמים לצורך ניבוי, סיווג נתונים ( Cluster Data), או לקבלת סדרת החלטות רצופות (Sequential Decisions) תוך אינטראקציה אופטימלית עם הסביבה.

ניתוח סטטיסטי עוסק באופן מסורתי ביצירת השערות (מבלי להסתכל על הנתונים) ולאחר מכן בבדיקת ההשערות באמצעות נתונים. למידת מכונה שונה מניתוח סטטיסטי בכך שהיא לא יוצרת השערות אלא גוזרת את המודל לחלוטין מתוך הנתונים.

היבט חשוב של למידת מכונה הינו תיקוף ובדיקה (Validation and Testing). רוצה לומר שיש לבדוק ולתקף מודלים שנוצרו באמצעו אלגוריתמים של למידת מכונה, בעזרת סט נתונים אחר, כזה שלא שימש ליצירת המודל (Out-of-Sample). מחד גיסא, מודל מורכב מדי עשוי להתאים את עצמו יתר על המידה (Over-fit, ללמוד יותר מדי טוב את) לנתונים ששימשו לאימון המודל ובכך הוא עלול שלא להצליח להכליל (Generalize) באופן מספק דיו את הנתונים החדשים. מאידך גיסא, מודל פשוט מדי עלול שלא להצליח לתפוס היבטים חשובים של הנתונים. למידת מכונה גורסת שיש לחלק את הנתונים הזמינים ל- 3 סטים של נתונים. סט האימון ( Training Set) משמש לגיבוש/פיתוח מודלים אלטרנטיביים. סט התיקוף (Validation Set) משמש לבדיקה עד כמה המודלים מכלילים טוב את הנתונים החדשים. סט הבדיקה (Testing Set) נשמר בצד לאורך כל התהליך שתואר עד כה ומשמש כמבחן סבירות סופי לרמת הדיוק של המודל הנבחר.

טרם השימוש באלגוריתם של למידת מכונה, חשוב מאוד לנקות תחילה את הנתונים. המאפיינים (Features, המשתנים המסבירים) המהווים את הנתונים יכולים להיות נומריים או קטגוריים. בכל מקרה עשויים להיות מצבים של חוסר עקביות (Inconsistencies) באופן שבו

הנתונים הוכנסו למאגר הנתונים. לפיכך, יש לזהות ולתקן מצבים של חוסר עקביות. חלק מהתצפיות עשויות להיות לא רלוונטיות למשימה הנוכחית ועל כן יש להשמיטן. בנוסף, יש לבדוק שאין תצפיות כפולות או כפילויות בנתונים, דבר שעלול ליצור הטיות. יש להשמיט חריגים אשר נוצרו בוודאות כתוצאה מטעויות הקלדה או מטעויות בהכנסת הנתונים למאגר. לבסוף, יש לטפל בנתונים חסרים באופן שלא יטה את התוצאות.

משפט בייס (Bayes Theorem, נוסחת בייס) הוא תוצאה המשמשת לעתים כאשר היא נדרש לכמת את אי הוודאות. משפט בייס הוא דרך להפוך משהו להתניה. נניח שאנו רוצים לדעת מהי ההסתברות שמאורע Y יתרחש ונניח שאנו גם יכולים לדעת האם מאורע אחר שקשור למאורע Y, נקרא לו מאורע X, התרחש או לא. עוד נניח שעל סמך ניסיון אנו יודעים את ההסתברות המותנה (Intensity) שמאורע X יתרחש בהינתן שידוע שמאורע Y התרחש. למעשה משפט בייס מאפשר לנו לחשב את ההסתברות המותנה שמאורע Y יתרחש בהינתן שידוע שמאורע X התרחש.

ללמידת מכונה ישנה טרמינולוגיה משלה אשר שונה מזו המסורתית המשמשת בסטטיסטיקה. במסגרת הטרמינולוגיה של למידת מכונה: מאפיין (Feature) הוא משתנה אשר לגביו יש לנו תצפיות; יעד (Target) הוא המשתנה אשר עליו אנו רוצים לבצע תחזיות; תוויות (Labels) הן תצפיות על היעד; למידה בהשגחה (Supervised Learning) היא תחום של למידת מכונה שבמסגרתה אנו משתמשים בנתונים על המאפיינים והיעדים לצורך ניבוי היעד על סמך נתונים חדשים; למידה ללא השגחה (Unsupervised Learning) היא תחום של למידת מכונה שבמסגרתה אנו מנסים למצוא דפוסים בנתונים על מנת לסייע לנו בהבנת מבנה הנתונים (בלמידה ללא השגחה אין יעד ועל כן אין גם תוויות); למידה בהשגחה למחצה (Semi-Supervised Learning) היא תחום של למידת מכונה שבמסגרתה אנו מבצעים תחזיות על היעד על סמך נתונים אשר לחלקם יש תוויות (קרי, יש להם ערכים של היעד) וליתר אין תוויות (קרי,

אין להם ערכים של היעד); למידה בחיזוקים (Reinforcement Learning) היא תחום של למידת מכונה שבמסגרתו אנו יוצרים אלגוריתמים לקבלת סדרת החלטות רצופות כאשר מקבל ההחלטה פועל בתוואי של סביבה משתנה.

**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.IL.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 סיכוני שוק, אשראי, תפעוליים, מודל, נזילות והשקעות לצורכי יישום הוראות רגולטוריות ותקינה חשבונאית, פיתוח ,יישום ותיקוף מודלים 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 Clustering,k-means Clustering, Density-based 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

Intrinsic Value
Independent Business Appraisers
Recognize. Measure. Present the Result.

Accuracy
IAVFA
לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל
Israel Association of Valuators and Financial Actuaries

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).