

## למידה בהשגחה (Supervised Learning): יער אקראי (Random Forest)

לעיתים, יותר מאלגוריתם אחד של למידת מכונה משמש לביצוע תחזיות, ואז למעשה ניתן לשלב את התוצאות. דבר זה מכונה שימוש בשיטות הרכב (Ensemble Methods). במקרה של סיווג, ניתן להשתמש בנוהל של הצבעה (Voting). אם מרבית האלוגריתמים מנבאים תוצאה ספציפית, הרי שניתן לבחור בה כתוצאה החזויה. במקרה של ניבוי ערך נומרי, ניתן למצע את התוצאות המתקבלות מהאלגוריתמים השונים על מנת לייצר תחזית מורכבת (Composite Prediction).

אלגוריתם למידת מכונה מסוג יער אקראי (Random Forest) נוצר באמצעות בניית עצים שונים רבים ושילוב התוצאות המתקבלות באופן שכרגע תיארנו. ניתן ליצור את העצים על ידי דגימה (Sampling) מתוך התצפיות או מתוך המאפיינים (או שניהם). לחילופין ניתן ליצור את העצים גם על ידי קביעה אקראית (Randomizing) של ערכי הסף.

Bagging הינו מונח המשמש כאשר תתי-קבוצות שונות של תצפיות בסט האימון משמשים ליצירת מודלים.

Boosting הינו ורסיה של שיטות הרכב שבה מודלים של ניבוי נבחרים ברציפות וכל מודל נועד לתקן את השגיאות של המודל הקודם. אחת הדרכי לסיווג היא להגדיל את המשקל של התצפיות שסווגו לא נכון (Misclassified). דרך אחרת היא להשתמש במכונת למידה לניבוי השגיאות הניתנות על ידי המודל הקודם.

## 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 סיכוני שוק, אשראי, תפעוליים, מודל, מזילות 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 וחזוי מצבים פיננסיים מורכבים וכן העברת סמינרי הדרכה והשתלמויות בתחומי התמחות: 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 ושלפה 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



ותוצרין, פיתוח כלים, מודלים, תהליכים ומערכות בתחום האנליזה, תוך שימוש בכלי אנליזה מתקדמים (EXCEL, VBA ו-R).