**AI**- התפיסה שמכונות יכולות ללמוד, להבין ולהסיק כמו בני אדם. כלומר ניתן להסתכל על התוכנה או המכונה ולומר שיש בה תבונה אנושית. **Machine Learning** - המערכת משפרת את הביצועים שלה יחד עם הניסיון.  **- Training set**אסופה של אימון, כלומר המידע שהמחשב מקבל, מתאמן ולומד עליו איך לנתח את המידע, כלומר סט האימון מגיע יחד עם "התשובה" (לדוגמה עבור כל רשומה אם זה גבר או אישה) וכך לומד באופן עצמאי איך להגיע לתשובה לבד לפי המידע שניתן לו. **Test set** – סט הבדיקה , שולחים למכונה סט שהוא לא ראה ובודקים האם הוא יודע לחזות את התשובה. סט האימון תמיד גדול מסט הבדיקה ורשומה שנמצאת בסט האימון לעולם לא תהיה בסט הבדיקה (מצבים כאלו יכולים לגרום לoverfitting).- **Validation set**לוקחים את סט הבדיקה ומפרקים אותו לואלידיישן סט ולטסט סט. **ההבדל בין אלגוריתם למודל**- מודל הוא ביטוי מתמטי הממיר קלטים לפלטים מסוימים, בניגוד לאלגוריתם, המודל בתהליך הלמידה לומד חשיבות של כל פרמטר, אלגוריתם יודע בתחילת הריצה מבחינה סטטיסטית משקל של כל פרמטר. **Data Science** - מדעי הנתונים, תחום שלוקח בחשבון שיטות מדעיות, תהליכים ואלגוריתמים ומנסה להוציא תובנות ממידע מסודר או לא מסודר (דאטה סיינס כולל בתוכו משין לרנינג). **Big data**  - התעסקות עם בסיסי נתונים מאוד גדולים.**Deep learning** – דורש מידע גדול מאוד (ביג דאטה), היתרון של למידה עמוקה – עובד יותר טוב מאשר משין לרנינג בעיקר כשיש הרבה דאטה, והחיסרון הוא שקשה לבין למה זה עובד. **Expert based feature**- תכונות לזיהוי אובייקטים למשל שיער פנים, גובה ועוד. **Object detection**  - זיהוי האובייקט. **Bounding box** – תיבה הסוגרת את גבולות האובייקט. - **Overfitting** התאמת יתר (Overfitting) היא בעיה יסודית ב[סטטיסטיקה](https://he.wikipedia.org/wiki/%D7%A1%D7%98%D7%98%D7%99%D7%A1%D7%98%D7%99%D7%A7%D7%94) וב[למידת מכונה](https://he.wikipedia.org/wiki/%D7%9C%D7%9E%D7%99%D7%93%D7%AA_%D7%9E%D7%9B%D7%95%D7%A0%D7%94) שבה המודל מותאם יתר על המידה לאוסף מסוים של נתונים (למשל האוסף שהיה זמין לשם אימונו) ועל כן מצליח פחות בביצוע תחזיות. התאמת יתר מתרחשת כאשר המודל נקבע על ידי יותר פרמטרים מאשר הנתונים מצדיקים. עודף הפרמטרים מאפשר למודל ללמוד את הרעש הסטטיסטי כאילו הוא מייצג התנהגות אמיתית. **Ground truth** – נתוני האמת. **אנטרופיה**- המטרה של מדד זה היא להעריך את אי-הוודאות של משתנה מקרי כלשהו (X). **משתנה בדיד**- ערכים שאי אפשר להשוות בגודליהם למשל גבר או אישה. **משתנה רציף-** משתנים שיש משמעות לגודלם למשל טמפרטורה, גובה משקל ועוד. - **בעיית קלסיפיקציה**- זיהוי קלאס על בעיות בדידות (גבר או אישה צבע ועוד). **בעיית רגרסיה**- זיהוי קלאס עם תכונות רציפות. **דיסקרטיזציה**- הפיכה משתנה רציף לבדיד. **קלסיפיקציה**- חיזוי קטגוריות (קלאסים). **Majority voting** – הרוב קובע. **ID3** – אלגוריתם לבנית עץ החלטה, לא מבטיח מניעת Overfitting בדף הבא יש דוגמה לאלגוריתם. **Preprocessing** – ניקוי המידע הלא רלוונטי. **הערה-** כאשר בונים עץ החלטה, ככל שהעץ עמוק יותר כך יש יותר סיכוי שהוא יסבול מOverfitting. **דרכים למניעת Overfitting בעצי החלטה- pre-pruning , גיזום מוקדם** – נרצה לעצור את בנית העץ, נקבע איזשהו סף וממנו נפסיק לפתח את קודקודי העץ, הבעיה בזה היא הקושי לבחור תנאי סף מתאים. **Post-pruning** – נבנה עץ ענק שנדע שהוא לא טוב וננסה לגזום אותו, בד"כ הגיזום יתחיל מהרמה הנמוכה ונעלה אט אט, איך ומתי נדע שהעץ הוא אופטימלי? בכל שלב אנו גוזמים את העץ נריץ את סט הבדיקה ונראה אם שיפרנו. **Cross validation** – מריצים כמה טסטים כאשר בכל פעם חלוקה אחרת היא טסט הבדיקה.  
**מבחן חי בריבוע**- אלגוריתם לבנית עץ החלטה המבטיח שהעץ לא יסבול מOverfitting . – **PEP - Pessimistic error pruning**עוד מבחן. במקרה כאן קודם נבנה את העץ ורק אחר כך נבצע גיזום. החצי שמחברים במשוואה היא עבור תיקון הרציפות. **אלגוריתם דיסקרטיזציה** – נבצע אם ישנם רצפים. לא נוכל עבור משתנה רציף לבנות עץ באמצעות ID3 או באמצעות pruning. **Gini index** – שיטה לפיצול, מחלקת את הקלאסים לפי הקלאס הגדול ביותר, טובה בד"כ להחלטות בינאריות (כן או לא, 0 או 1). **Towing**  - שיטה לפיצול, מחלקת את הקלאסים חצי חצי לפי גודלם (כמות הערכים בדאטה סט) – דומה לאנטרופיה אבל יותר גמישה, טובה למקרים בהם יש הרבה קלאסים. **הערה-** מבחינת חוזק השיטות- מהנמוך לגבוה, שיטות בייסיאניות, עץ החלטה, אנסמבל ורשתות נוירונים. **Ensemble methods** - מחלקים את סט האימון לכמה פיסות מידע ועל כל פיסת מידע מפעילים אלגוריתם שונה. היתרונות- משפר ביצועים, קל למימוש. חסרונות- תופס הרבה מקום, סוג של קופסה שחורה. **שלושה סוגים של אנסמבל מטודס: - Bagging** כרוכה בשימוש באותו אלגוריתם אך באימונו על דגימות אקראיות שונות של סט האימון (training set) או של המאפיינים. **Boosting** מתייחס לשיטה מורכבת שבמסגרתה שיטות החיזוי מיושמות ברצף (אחת אחרי השנייה), כאשר כל אחת מנסה לתקן את השגיאות של קודמתה, הליך זה מוכר בשם **AdaBoost. יער אקראי**- כפי ששמו מרמז עליו, מורכב איך לא מעצי החלטה. העצים נוצרים לרוב על ידי דגימה מתוך המאפיינים או מתוך התצפיות (כלומר, באמצעות גישת ה- bagging). כל אחד מהעצים נותן תוצאה לא-אופטימאלית (Suboptimal) אך באופן כללי, על פי רוב החיזוי בדרך זו משתפר. **שיטת בייס** – שיטה לסיווג קלאסים לפי נוסחאת בייס, היתרונות- פשוטה למימוש, חסרונות- התכונות לא תלויות אחת בשניה כאשר יש מקרים בהם כן קיים תלות.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי אנטרופיה

לאחר מכן נעשה את אותו הדבר עבור Parents ועבור Money והערך הכי גדול מבינהם יהיה המאפיין שלפיו נפצל את העץ הפעם הראשונה. לאחר מכן נרצה לפצל את העץ שוב, אז נעבור על התהליך שוב רק שלא נחשב את המאפיין שכבר השתמשנו בו לפיצול.

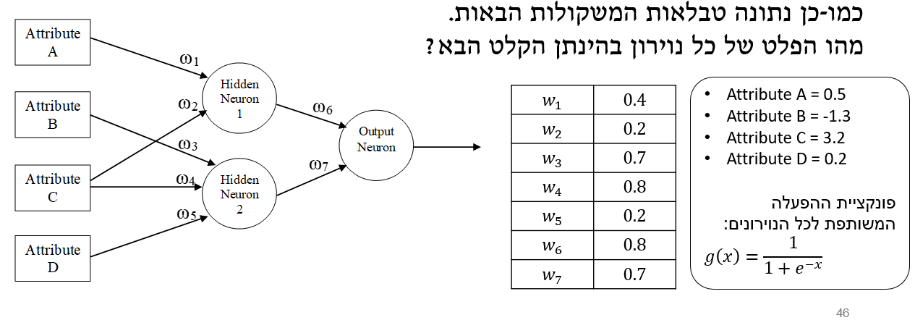
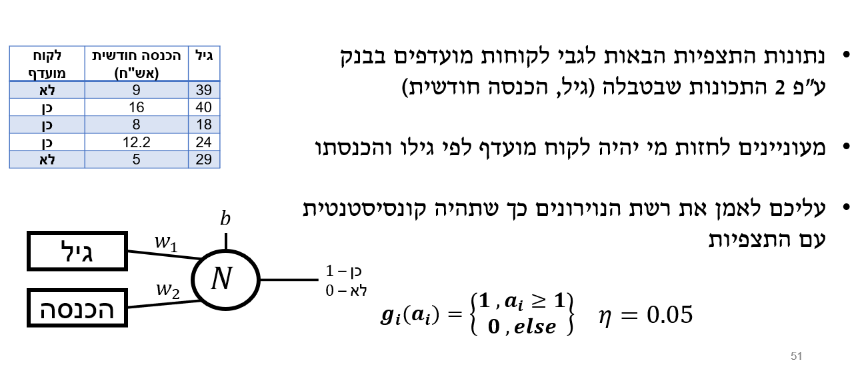
ID3

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

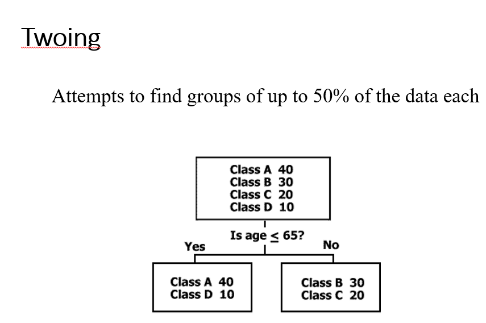
התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

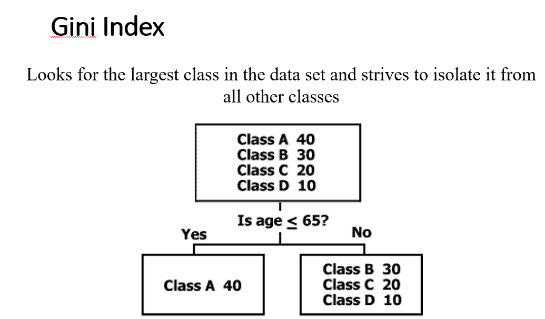
התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיתמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי