

Advanced Data Science Techniques

130 שעות לימוד אקדמיות

תיאור התפקיד:

Data Scientist אחראי על איסוף, ניתוח והגשת התוצאות מניתוח כמויות גדולות או קטנות יותר של נתונים. תהליך זה משמש לקבלת החלטה חשובה עבור ארגון, אשר יכולה להשפיע על הצמיחה עסקית ולעזור להתמודד עם התחרות בשוק. תפקידו של Data Scientist הינו ניתוח נתונים במטרה לחלץ תובנה מושכלת מהתוצאות. על מנת לתת מענה מקצועי לצורך עסקי, מדען נתונים חייב להיות מיומן בתחומים עיקריים:

- חשיבה ויכולות סטטיסטיות/מתמטיות
- בעל יכולות פיתוח/תכנות ברמה גבוהה
- מיומנות להגדיר אוסף נתונים ומשתנים נכונים
- יכולת לזהות את הבעיות המאתגרות ביותר בניתוח נתונים
- יכולת להתמודד עם אוסף של מערכי נתונים לא מובנים וגם נתונים מובנים, ממקורות שונים
- בנייה ויישום מודלים ואלגוריתמים - מודלים של Machine Learning
- תקשור ממצאים לבעלי העניין באמצעות כלי ויזואליזציה
- בעל היכרות מעמיקה עם הביזנס ובעל יכולות תקשורת והצגה גבוהות
- יכולת להוביל פרויקט Data Science מקצה לקצה

כיום לא מספיק לדעת להפעיל וליישם מודלים על דאטה, ובפועל נדרשות מיומנויות ברמה גבוהה יותר על מנת לענות על הדרישות שתיארנו למעלה. יכולות אלו נרכשות עם הזמן ועם הנסיון, וכן עם תהליכי למידה של נושאים משלימים, מתקדמים והיכרות מעשית עם טכניקות שונות בעבודה על נתונים.

תיאור ההכשרה:

המתודולוגיה של פרויקט Data Science ברורה (CRISP), אבל בכל אחד מהשלבים של תהליך Data Science יש הרבה עומק שמגיע מיכולת הסקה סטטיסטית, היכרות עם טכניקות מתקדמות ושימוש במודלים בצורה מושכלת. ככל שמדען נתונים מכיר יותר טוב ויותר לעומק את המודלים שיכול להשתמש בהם – התהליך שהוא מבצע הרבה יותר מקצועי.

מטרת המסלול היא להעניק ידע מתקדם למדעני נתונים מתחילים, שעברו הכשרות שונות ונכנסו לתפקיד בתעשייה בשנה האחרונה. וכן להעמיק בנושאים מתקדמים מתחום Deep Learning, ובעיקר בנושאי עיבוד תמונה וניתוח טקסטים.

מתכונת הקורס מבוססת על תהליך למידה דרך שימוש ב-Usecases רלוונטיים מעולמות תוכן שונים. במסגרת כל Usecase יצופו אתגרים חדשים, ואנו נציג רעיונות ועקרונות מתקדמים להתמודדות עימם. הורה שבה הקורס מתנהל הינה יישומית ופרקטית, תוך כדי הסברים ותרגול מעשי רב. במהלך הדרכה ושימוש ב-Usecase נקיים דיון מקצועי, נדבר על פתרונות אפשריים ונעלה רעיונות יצירתיים למימוש. מסלול הכשרה זה מלווה בתרגול מעשי במהלך השיעורים, במסגרתו יוכלו לממש את הידע הנרכש במהלך השיעורים.

קהל יעד:

בעלי ידע ונסיון של כשנה בתחום Data Science

דרישות קדם:

- יעוץ עם גורם מקצועי
- תכנות בפיתון ברמה גבוהה כולל Pandas ו-ScikitLearn
- בעלי ידע ונסיון ב-Data Scientist Junior של שנה בתעשייה:
 - מעבר מבחן כניסה הבוחן את רמת הידע
 - במפגש השני של הקורס יידרש כל סטודנט להציג פרויקט שלו מהעבר

תוכנית הלימודים:

Recommendation engines

במודול זה נביר סוגים שונים של מנגנוני המלצה, ואלגוריתמים שונים לממושם. נתמודד עם בעיית הדלילות בדאטה – Sparsity, נביר התמרות (טרנספורמציות) שימושיות.

- Content based/Collaborative filtering
- Dealing with Sparse Data
- Dimensionality reduction
- Matrix factorization

Deep learning basics

במודול זה נערוך היכרות עם מושגי היסוד של עולם ה-Deep Learning. נביר את הארכיטקטורות הבסיסיות ואת מנגנוני הרגולריזציה (ריסון) המרכזיים. כמו כן נעמיק את ההבנה באופן פעילותו של תהליך האופטימיזציה.

- Neural networks & MLP model
- Deep architectures with Keras
- Main Concepts (batch normalization, pooling, attention, optimizers, etc.)

Deep Learning Applications

במודול הזה נביר את היישומים המרכזיים והבולטים ביותר של עולם ה-Deep Learning, עיבוד תמונה וניתוח טקסט. עולם המושגים שנעניק במודול הזה יאפשר לסטודנטים להיכנס לאחד התחומים האלה בהמשך ולהתמחות בו.

- Image processing basics
 - Classic terminology & tools
 - Convolutional Neural Networks (CNNs)
- NLP basics
 - Classic terminology
 - Text mining tools
 - Recurrent Neural Networks (RNNs)

Advanced ML concepts

במהלך הקורס ניחשף ונתרגל עקרונות וטכניקות מתקדמים בתחום Machine Learning. מדובר בטכניקות רחבות השימושיות בשלבים שונים של תהליך Data Science ויישומן בתהליכי פיתוח מודל עשוי להיות קריטי.

- Advanced cross-validation
- Error Analysis
- Dimensionality reduction
- Anomaly/outlier detection
- Feature engineering techniques
- Statistics