



Certified Data Scientist

לימודי תעודה להכשרת מדעני נתונים

עם התמחות ב Generative AI

בית הספר
להייטק וסייבר
המערך לתוכניות ייעודיות
אוניברסיטת בר-אילן



אוניברסיטת בר-אילן, המערך לתוכניות ייעודיות
המדור לתוכניות מובנות ולימודי תעודה
03-7384481 | טלפון | Desigprog.biu.ac.il

* לימודים לא אקדמיים

Certified Data Scientist

המסלול השלם להכשרת מדעני נתונים

עם התמחות ב Generative AI

מנהל אקדמי: מר ערן שחם, מנהל בית הספר להייטק וסייבר באוניברסיטת בר-אילן

יועצים אקדמיים: ד"ר תומס קרפטי ומר אביעד רוזנהק

מדעי הנתונים משתנה במהירות ותחום ה Generative AI הוא אחד המרגשים. מדעני נתונים עם התמחות ב Generative AI, מבוקשים מאוד בתעשייה ויש להם פוטנציאל להשפיע רבות על העולם. ה Data Scientist מבצע מחקרים להפקת תובנות עסקיות לארגון (Business Intelligence), מטייב ומסדר את המידע המשמש למחקרים, מפעיל אלגוריתמים שונים של Machine Learning, כריית מידע ומידול ובכך מסייע בבניית תהליכי הכנת המידע ואופטימיזציה של האלגוריתמים השונים. תוך שילוב כלים סטטיסטיים, הוא מסוגל לחזות את העתיד בהינתן מידע קיים.

ל- Data Scientist ידע רב בנושאים הבאים:

- אינטגרציה - איסוף מידע ממגוון מערכות ועבודה עם כמויות גדולות של מידע (Big Data) ועיבוד מידע לא מובנה (Unstructured).
- חקירה - תכנות וניתוח סטטיסטי, יצירת חיבור בין בסיסי נתונים שונים.
- ניתוח אנליטי - חיזוי, כריית מידע, אופטימיזציה, עיבוד מידע טקסטואלי ואנליזה לנתונים.
- הצגה - פרסום תוצאות על בסיס ניתוח המידע שנאסף.

Generative AI היא תת-תחום של בינה מלאכותית שמתמקדת בפיתוח אלגוריתמים, המסוגלים ליצור נתונים חדשים. זה יכול להיות טקסט, קוד, תמונות, צלילים, או כל סוג אחר של דאטה. קיימות מספר סיבות מרכזיות ללימוד התחום:

- **צמיחה חסרת תקדים:** עם הקמתה, ChatGPT היה המוצר הצומח ביותר אי פעם, מה שמדגיש את הפוטנציאל העצום ואת העניין בטכנולוגיות בינה מלאכותית.
- **מהפכה בפרודוקטיביות:** כמעט כל הארגונים משלבים כיום בינה מלאכותית גנרטיבית באסטרטגיות החדשנות שלהם. השינוי הטכנולוגי הזה מגדיר מחדש את האופן שבו עסקים מתחדשים ומתחרים.
- **מיומנויות מבוקשות מאוד:** יש ביקוש הולך וגובר לאנשי מקצוע בעלי מומחיות בהיבטים שונים של AI גנרטיבי. זה כולל מיומנויות בהנדסת פרומפטס, היכרות עם מסגרות כמו AutoML ו-LangChain, והיכולת לעשות אימון משני למודלי בסיס כדי לתת מענה לדרישות עסקיות ספציפיות.
- **בינה מלאכותית כללית:** ChatGPT השיגה את אבן הדרך של עמידה במבחן טיורינג, מדד ליכולת של מכונה להפגין אינטליגנציה דמוית אדם. הישג זה האיץ את החיפוש של התעשייה לבנות בינה כללית מלאכותית (AGI) על בסיס מודלי שפה גדולים (LLMs).

דוגמאות לשימוש ב Generative AI:

- **פרודוקטיביות ואוטומציה משופרת:** ייעול משימות תכנות, מדעי נתונים ושירות לקוחות עם יכולות כמו השלמה אוטומטית של קוד, המרת מפרטים לתפוקות מוחשיות וייזום תהליכים באופן אוטונומי.
- **מהפכה בתעשיות יצירתיות:** דמוקרטיזציה של אמנות ויצירת תוכן, תמונות ומוזיקה בשפה אנושית.
- **מהפכה במדע ובמחקר:** פריצות דרך בדיסציפלינות כמו מתמטיקה, ביולוגיה, רפואה וחקלאות על ידי שימוש בבינה מלאכותית גנרטיבית ליצירת טיפולים רפואיים חדשים, הנדסת חלבונים, תכנון ארכיטקטורות טכנולוגיות חדשניות כמו מעבדים, וביצוע מחקר באופן אוטונומי, כולל הוכחות מתמטיות והפקת השערות.

בתוכנית להכשרת מדעני נתונים המתמחים ב Generative AI נכיר את כל אלו. בסיום הלימודים יגישו התלמידים עבודת גמר המתבססת על מחקר נתונים תוך שימוש בטכנולוגיות החדשות ביותר. תיק העבודות אותו יוצר התלמיד בקורס ניתן להצגה מול מעסיקים בתהליך הקבלה למשרות בתחום זה.

מטרת התוכנית

התכנית תכין את התלמיד להיות מדען נתונים מיומן עם יכולת עבודה עצמאית וכחלק מצוות מולטי דיסציפלינרי. התכנית תספק את הידע ואת הכלים הדרושים כדי לפתח פרויקטים של למידת מכונה (Machine Learning) וליישם אותם במערכות הייצור. עם סיום התכנית התלמידים יוכלו:

- לעבוד על פרויקט מדעי נתונים מתחילתו ועד סופו.
- לעבוד כצוות משימה בסביבה משותפת.
- לכתוב פרוטוקול מחקר לפני תחילת הפרויקט.
- לספק נתונים נקיים ומכנים לאנליזה.
- לפתח מודלים מנבאים בשימוש כלים מתאימים עבור הבעיה ולפי הדרישות של הלקוח.
- לכתוב דו"ח סיכום המסביר את המתודולוגיה ששימשה את הפרויקט והמודלים שפותחו יחד עם תיקוף המודלים.
- הטמעת המודלים שפותחו והדרכה של משתמשי הקצה לעשות שימוש נכון ומושכל בהם.

יתרונותיה הבולטים של התוכנית

- מיקוד בכלים הנפוצים בתחום בשפת הפיתוח **Python**: Sklearn, Keras, LangChain
- תעודה יוקרתית של אוניברסיטת בר אילן, ביה"ס להייטק וסייבר
- קורס מעשי במהותו, המשלב יישום תרגולים רבים ובניית פרויקט אישי במקביל לידע התיאורטי הנרכש.
- הקורס מועבר בהדרכתם של מרצים מובילים וותיקים המאשרים ע"י אוניברסיטת בר אילן.
- סדנאות מיתוג אישי ובניית פרזנטציה מנצחת לצורך מיקסום אפשרויות תעסוקה
- ליווי אישי של התלמיד לאורך כל המסלול וגם לאחריו.

אודות המנהל האקדמי, מר ערן שחם

- בעל נסיון של 30 שנה כמרצה ומנהל הדרכות בתחום הטכנולוגי והניהולי.
- מנהל ביה"ס להייטק וסייבר באוניברסיטת בר-אילן
- הקים וניהל את קורסי התעודה במחשבים בביה"ס ללימודי המשך של הטכניון במשך 8 שנים
- נהל את מרכז ההדרכה המורשה של Check Point במשך 18 שנה
- הקים ונהל את "מלם-תים הדרכה", מקבוצת מלם-תים, האחראית להכשרת 4,000 עובדי הקבוצה.
- נהל את קהילת אבטחת המידע במיקרוסופט ישראל

אודות היועץ האקדמי, ד"ר תומאס קרפטי

- רופא, מדען נתונים ומרצה בעל נסיון של 20 שנה
- בעל נסיון רב בפיתוח Full Stack בתחומי Data Warehouse ו Data mining בעולמות הרפואה
- מרצה 5 שנים בתחומי ה Data Science בטכניון ובאוניברסיטת בר-אילן
- מנהל את המעבדה הרפואית האנליטית המולטידיסציפלינרית ב HIT
- מרצה את תחומי ה Machine Learning בקורס

אודות היועץ האקדמי, מר אביעד רוזנהק

- מוסמך האוניברסיטה העברית למדעי המחשב
- מנהל מוצר ב YouTube, אשר השיק מוצר של אנליזה סטטיסטית
- הקים את livv.ai, מנהל מוצר ב comet.ml חברה בתחום MLOps
- מרצה בתחומי ניהול מוצר מבוסס דאטה, מדעי הנתונים, data engineering ופיתוח מתקדם בפיתוח.
- מעביר את תכני ה Generative AI בקורס



תנאי קדם

- תואר ראשון או רקע אקדמי בתחומי מערכות מידע/תעשייה וניהול/מדעי המחשב/כלכלה/תחומי המדעים
- ראיון קבלה ע"י יועץ אקדמי

מתכונת לימודים

- פתיחת הלימודים: 13.3.2024
- משך הלימודים: כ- 7 חודשים ויתקיימו בימים ראשון ורביעי בין השעות 17:30-21:30
- תכנית הלימודים כוללת 275 שעות אקדמיות.

שכר לימוד

- דמי רישום: 480 ₪.
- שכר לימוד: 14,500 ₪.
- חניה: עבור תשלום נוסף של 200 ₪ ניתן לחנות את הרכב באוניברסיטה בשעות הלימוד עד סוף ספטמבר.

מקום הלימוד

- הלימודים נערכים בבית הספר להייטק וסייבר, קמפוס אוניברסיטת בר-אילן, רמת גן במקביל מתבצעים באופן מקוון. כל השיעורים מוקלטים וזמינים לתלמידים במהלך הקורס ובסיומו.

זכאות לתעודה

- חובת נוכחות ב-80% מהמפגשים, ועמידה במטלות התכנית.
- לעומדים בדרישות התכנית תוענק תעודה מטעם אוניברסיטת בר אילן, המערך לתוכניות ייעודיות.

הערות

- פתיחת כל תכנית מותנית במספר הנרשמים.
- דמי ההרשמה אינם כלולים בשכר הלימוד ואינם מוחזרים, אלא במקרה של אי פתיחת התכנית על ידי היחידה, בכפוף לתקנון לימודים של לימודי התעודה באוניברסיטת בר אילן.
- היחידה מביאה לידיעת הנרשמים כי ייתכנו שינויים במערך התכנית, במועדי הלימודים והבחינות או בכל נושא אחר. הודעה על כל שינוי תימסר למשתתפים, בכפוף לתקנון לימודי התעודה באוניברסיטת בר אילן.

נהלי רישום

ההרשמה תבצע באמצעות חברת ערן שחם בע"מ, עבור אוניברסיטת בר אילן.

לפרטים והרשמה נא לפנות אל:

אסתר אסולין

052-5886002

esther.asulin@biu.ac.il



תכנית הלימודים

מודול 1 - מבוא למדעי הנתונים (10 ש"א):

- מדען הנתונים מנקודת ראותו של ה CDO
- את ההרצאה יעביר מנהל דאטה ארגוני ויסביר את הפרמטרים העיקריים באמצעותם נמדד מדען הנתונים
- מהו מדע הנתונים
- מחקר מדעי - שיטות מחקר
- אנטומיה של פרויקט במדעי הנתונים

מודול 2 - בסיס נתונים ותכנות SQL (30 ש"א):

- מבוא ל-SQL
- חיבור למסדי נתונים, יצירה ומחיקה של טבלאות, פקודות: SELECT, INSERT, DELETE, UPDATE
- תכנות SQL
- חיבור טבלאות JOINS: INNER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN, UNION
- שינוי טבלה ALTER, יצירת אינדקסים, טבלאות זמניות
- תכנות SQL מתקדם
- סינון ומיון נתונים, משתנים מחושבים: טרנספורמציות מתמטיות, פעולות על מחרוזת, פעולות על התאריך והשעה. אגרגציות של נתונים, פיבוטים PIVOT

מודול 3 - תכנות בסיסי ב-Python (20 ש"א):

- מבוא ל-Python, אנקונדה ומחברות Jupyter, תחביר Python, פונקציות, סוגי נתונים ב-Python ואופרטורים (מספריים, מחרוזות, רשימות, Tuples, Arrays, תאריך ושעה), שימוש בפונקציות ומתודות של אובייקטים
- SLICING, INDEXING AND REVERSE INDEXING
- לולאות, תנאים וערכים בוליאניים. שימוש ב Range

מודול 4 - Python למדעי נתונים (40 ש"א):

- עבודה עם נתונים באמצעות PANDAS
- גרפים עם ספריות Seaborn, Matplot, Pandas
- קלט פלט עם Pandas
- כתיבת פונקציות ופונקציות LAMBDA
- List comprehensions
- מבוא לניהול קבצים באמצעות Path
- מבוא ל TYPE HINTS
- מבוא ל OOP עם DATA CLASSES

מודול 5 - כלים לעבודה במדעי הנתונים (5 ש"א):

- כלים לעבודה נכונה ותקינה במדעי הנתונים המאפשרת שיתוף ושיחזור של פרויקטים: שמירת גרסאות git ו-github.
- שמירת תקינות סביבת העבודה: סביבות וירטואליות באמצעות Poetry

מודול 6 - נתונים (5 ש"א):

סולמות נתונים:

- סולם נומינלי
- סולם סודר
- סולם רציף
- סדרות עיתיות

סוגי נתונים:

- מחרוזות (Char & Varchars)
- מספרים שלמים (Integers)
- מספרים עשרוניים (Decimal)
- Double & Float
- תאריך ושעה
- בוליאניים
- טקסט
- ביטים

מקורות מידע:

- פורמטים של קבצים: טקסט CSV, TAB, Delimited, Json, אקסל, SAS, SPSS.
- HTML / XML: שפת Markup, סכימות, Parsing, תקני העברת נתונים בתעשייה (MISMO, FpML, NDC, OTA, HL7).
- בסיסי נתונים: רלציונליים (MS SQL Server, MySQL, PostgreSQL, Oracle, DB2, Access), מבוססות עמודה (MonetDB, C-Store, Teradata, Vectorwise / Paracel, Sybase IQ, InfiniDB, Vertica), מבוססות סדרות עתיות (eXtremeDB, גרפיט, InfluxDB), גרפיות-Graph-db.
- Non-SQL (Big Data): מגנון לאחסון ואחזור נתונים, אחסון ערך-מפתח (Key-value), אחסון מסמכים, אחסון גרפים, אחסון אובייקטים.

מודול 7 - מבוא לסטטיסטיקה עם Python (25 ש"א):

סטטיסטיקה:

- מבוא לסטטיסטיקה
- סוגי התפלגויות נתונים
- מדדי מרכז ושונות
- שיטות פרמטריות ולא פרמטריות
- מדדי הקשר
- מדדי הבדל

הסתברויות:

- משפט בייסיאנית והסתברות מותנית
- הסתברות פשוטה
- Prior distribution
- הסתברות עם תוצאות ספירות
- אקראיות והסתברות:
- התפלגות הבינומיאלית, הסתברות Bernoulli, הסתברות נורמלית
- הסתברות לאירועים תלויים ועצמאיים
- סימולציות מונטה-קרלו
- מודלים שרשרת מרקוב



מודלים מתמטיים:

- יצירת נתונים סינתטיים וסימולציות
- מודלי רגרסיה: רגרסיה לינארית
- מודלי קלסיפיקציה: רגרסיה לוגיסטית
- רגרסיות תלויות זמן: ניתוח הישרדות
- ניתוח סדרות עתיות: ARIMA

מודול 8 - למידת מכונה Machine Learning (65 ש"א):

הכנת נתונים:

- ייבוא נתונים ממקורות שונים וחיבור לדאטסט סופי
- תחקור נתונים (EDA: Exploratory Data Analysis)
- נורמליזציה
- ניקוי הדאטה
- זיהוי וטיפול בערכי קיצון
- זיהוי המנגנון של יצירת ערכים חסרים ושיטות לאימפוטציה
- שימוש ב Sklearn Pipelines

למידה ללא השגחה Unsupervised:

- ניתוח אשכולות (Cluster analysis)
- אשכולות K-Centroid (ממוצעים, K-medoids וכו')
- אשכולות מבוססי צפיפות (DBSCAN)
- אשכולות היררכיים

אסוציאציות:

- כללי אסוציאציות
- אלגוריתם Apriori

למידה בהשגחה- סיווג לעומת רגרסיה:

- מודלים לינאריים
- מודלים לוגיסטיים
- עצי החלטה ואלגוריתמים ensemble
- מכונת וקטורי תמיכה (SVM)
- KNN
- סדרות עתיות (Time series analysis)

תיכנון ולמידה של מודלי חיזוי:

- הנדסת משתנים
- בחירת משתנים
- חלוקת הנתונים - Cross-validation
- בחירת שיטת הערכה המתאימה ביותר (metrics)
- בחירת אלגוריתם מכונת למידה
- שיפור הניבוי: כוונן היפרפרמטרים, שיטות הרכבה ensemble.

מודול 9 - Deep learning - רשתות עצביות (25 ש"א):

רשתות עמוקות:

- הקדמה - מה היא למידה עמוקה? היסטוריה של התחום, האוניברסליות של רשת למידה עמוקה.
- רשתות עצביות עם השגחה Supervised: רשתות Feedforward, Gradient descent, מעדכנים, פונקציות הפעלה activation functions, פונקציות הפסד loss function.
- בניית ואימון רשת פשוטה באמצעות KERAS
- רשתות קונוולוציונליות convolutional networks
- שימוש ב FOUNDATIONAL MODELS, Transfer Learning, Fine tuning
- Embedding ו- אוטו- אנקודרים Autoencoders
- רשתות חוזרות ו-LSTMs

מבוא לעיבוד טקסט קלאסי:

- כלים קלאסיים בעיבוד טקסט - טוקניזציה, Text Statistics, Sentiment Analysis, מילות עצירה, Summarization, שימוש בספריית NLTK

מודול 10 - Generative AI (15 ש"א):

- מהו Large Language Model (LLM), כיצד הוא מאומן וכיצד הוא עובד
- עיבוד טקסט מודרני עם ChatGPT: Summarization, Sentiment Analysis, Translation, Classification
- מבוא לשימוש ב ChatGPT API באמצעות LangChain
- מבוא ל- Diffusion Models - TEXT TO IMAGE ושימושים

מודול 11 - פרקטיקום ופרויקט מסכם (25 ש"א):

פרויקט הגמר - פרוטוקול מחקר, תחקור נתונים וניקוי נתונים:

- חזרה: הנדסה ובחירת משתנים, יצירת מודלי חיזוי והערכתם
- QA של תהליך מכונת למידה (ML)
- שימוש ב EXPERIMENTATION MANAGEMENT כגון BIASES&WEIGHTS, דו"ח מחקר
- יישום מודל החיזוי במערכת Production באמצעות FLASK REST API
- ליווי כיתתי ופרטני פרויקט גמר והצגתו מול צוות המרצים

מיתוג אישי בשוק התעסוקה המשתנה (10 ש"א):

- הסבר על שוק התעסוקה ועל הפלטפורמות הדיגיטליות הרלוונטיות למיתוג אישי
- שימוש בלינקדין ככלי למציאת עבודה ומיתוג אישי
- דגשים לכתיבת קורות חיים ומערכות אוטומטיות לסינון מתעניינים ATS
- בניית תוכנית פיתוח והכשרה מקצועית

** המערך לתוכניות ייעודיות שומר לעצמו את הזכות לערוך שינויים בתוכנית הלימודים

הערה: מדהים שאת ההקדמה, כולל תמונת השער, יצרנו באמצעות הטכנולוגיה

הזו תוך דקות. חישבו איך הקורס יכול למנף את הקריירה שלכם.