



Certified Data Scientist

לימודי תעודה להכשרת מדעני נתונים

עם התמחות ב Generative AI

בית הספר
להייטק וסייבר
המערך לתוכניות ייעודיות
אוניברסיטת בר-אילן





Certified Data Scientist

המסלול השלם להכשרת מדעני נתונים

עם התמחות ב- Generative AI

מנהל התוכנית: מר ערן שחם, מנהל ביה"ס להייטק וסייבר באוניברסיטת בר-אילן
יועצים מקצועיים: מר אריק ווסרמן

תחום מדעי הנתונים משתנה במהירות בעולם עתיר הדאטה בו אנו חיים, ואילו תחום ה- Generative AI הוא אחד התחומים המרגשים עם פיתוחים חדשים ומעוררי דמיון. מדעני נתונים עם התמחות ב- Generative AI מבוקשים מאוד בתעשייה, ויש להם פוטנציאל להשפיע רבות על הארגון בו הם פועלים. ה- Data Scientist מבצע מחקרים לצורך הפקת תובנות עסקיות עבור הארגון (Business Intelligence), מטייב ומסדר את המידע המשמש למחקרים, מפעיל אלגוריתמים שונים של Machine Learning ומנהל כריית מידע ומידול. בכך הוא מסייע בבניית תהליכים של הכנת המידע ואופטימיזציה של האלגוריתמים השונים. תוך שילוב של כלים סטטיסטיים, ה- Data Scientist מסוגל לספק תחזית לעתיד בהינתן המידע הקיים.

ל- Data Scientist ידע רב בנושאים הבאים:

- אינטגרציה - איסוף מידע ממגוון מערכות, עבודה 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 כגון טקסט, קוד, תמונות, צלילים, או כל סוג אחר של דאטה. קיימות מספר סיבות מרכזיות ללימוד התחום:

- צמיחה חסרת תקדים: עם הפצתו, ChatGPT הפך למוצר הצומח ביותר אי פעם, כהוכחה לפוטנציאל העצום של התחום ולעניין בטכנולוגיות של 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, כולל מיומנויות בהנדסת פרומפט, היכרות עם מסגרות כמו AutoML ו-LangChain, והיכולת לעשות אימון משני למודלי בסיס כדי לתת מענה לדרישות עסקיות ספציפיות.
- בינה מלאכותית כללית: ChatGPT השיג 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 של תכנות, מדעי נתונים ושירות לקוחות עם יכולות כמו השלמה אוטומטית של קוד, המרת מפרטים לתפוקות מוחשיות וייזום תהליכים באופן אוטונומי
- מהפכה בתעשיות יצרניות: דמוקרטיזציה של אמנות ויצירת תוכן, תמונות ומוזיקה בשפה אנושית.

מהפכה במדע ובמחקר: פריצות דרך בדיסציפלינות כמו מתמטיקה, ביולוגיה, רפואה וחקלאות על ידי שימוש בבנייה 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

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

- לעבוד על פרויקט של מדעי נתונים מתחילתו ועד סופו באופן עצמאי.
- לעבוד כצוות משימה בסביבה משותפת.
- לכתוב פרוטוקול מחקר לפני תחילת הפרויקט.
- לספק נתונים נקיים ומוכנים לאנליזה.
- לפתח מודלים מנבאים תוך שימוש בכלים 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 ביה"ס להייטק וסייבר אוניברסיטת בראילן.
- קורס מעשי במהותו, המשלב יישום באמצעות תרגולים רבים ובניית פרויקט אישי, כל זאת במקביל לידע התיאורטי הנרכש.
- הקורס מועבר בהדרכתם של מרצים מובילים וותיקים מהתעשייה, המאושרים ע"י אוניברסיטת בראילן.
- סדנאות מיתוג אישי ובניית פרזנטציה מנצחת לצורך מיקסום אפשרויות תעסוקה.
- ליווי אישי של התלמיד לאורך כל המסלול וגם לאחריו.

אודות מנהל התוכנית ומנהל בית הספר להייטק וסייבר, מר ערן שחם

- בעל ניסיון עשיר של למעלה מ-30 שנים כמרצה ומנהל הדרכות בתחום הסייבר ואבטחת המידע.
- מנהל בית הספר להייטק וסייבר של אוניברסיטת בראילן.
- מנהל מרכזי הדרכה מורשים של החברות המובילות בעולם בתחום הווירטואליזציה, תקשורת ואבטחת מידע Amazon AWS, VMware, EC-Council, Check Point, Forcepoint, ISC2, ומכשיר אינטגרטורים בקורסים המתקדמים.
- מכשיר לקוחות ושותפים של Check Point בטכנולוגיות מתקדמות כ-20 שנה.
- יזם, הקים וניהל את קורסי התעודה במחשבים ביחידה ללימודי המשך של "מוסד הטכניון" במשך 8 שנים, בת"א, חיפה וירושלים.
- הקים וניהל את "מלם-תים הדרכה", מקבוצת "מלם-תים", האחראית להכשרת 4,000 עובדי הקבוצה.
- ניהל את קהילת אבטחת המידע במיקרוסופט ישראל.

אודות היועץ המקצועי, מר אריק ווסרמן

- מומחה בתחום תשתיות סייבר ו-AI.
- דוקטורנט בתחום ה-AI באוניברסיטת בר-אילן.
- פרסם עבודות מחקר בתחום, שלהן השפעה בסייבר ובתחום הבינה המלאכותית.

אודות המרצה, מר עופר זוין

- מומחה והתמקצעות באקדמיה ובתעשייה בתחום Deep Learning Time Series.
- בוגר תואר ראשון + שני בכלכלה + מחקר פיתוח אלגוריתמים מטעם אוניברסיטת בר-אילן ובוגר הטכניון בתחום Data science.
- בוגר תואר שני בתחום Clark University – Data Information.
- בוגר תואר שלישי בתחום ה-Data science מטעם אוניברסיטת בן גוריון.
- מנהל מעבדת פינטק + ML & DS באקדמיה בתחום Data Sciences.
- מנהל תחום מדען נתונים בשוק ההון באחד מוסדות פיננסים גדולים.
- מרצה בתחום באוניברסיטת בן גוריון ובמכללות נוספות.

תנאי קדם

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

מתכונת לימודים

- פתיחת הלימודים: 08.01.2025
- משך הלימודים: כשבעה חודשים, בימים ראשון ורביעי בין השעות 17:30 – 21:30.
- * שיעורי השלמה יתכנו בימים שאינם ימי הלימוד הרגילים.
- תוכנית הלימודים כוללת 275 שעות אקדמיות פרונטליות.

שכר לימוד

- דמי רישום: 480 ₪.
- שכר לימוד: 14,500 ₪.

מקום הלימוד

- הלימודים נערכים בבית הספר להייטק וסייבר, קמפוס אוניברסיטת בר-אילן, רמת גן ובמקביל באופן מקוון.
- כל השיעורים מוקלטים וזמינים לתלמידים במהלך הקורס ובסיומו. בכל הכיתות קיימות מערכות מולטי מדיה מתקדמות, המאפשרות לסטודנטים לשאול שאלות ולשמע את כל שאר הסטודנטים.
- בהתאם להוראות פיקוד העורף והוראות האוניברסיטה, ייתכן שחלק מהשיעורים יועברו באמצעות זום.

זכאות לתעודה

- חובת נוכחות ב-80% מהמפגשים, ועמידה במטלות התוכנית.
- לעומדים בדרישות התוכנית תוענק תעודה מטעם ביה"ס להייטק וסייבר של אוניברסיטת בר-אילן.

הערות

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

נהלי רישום

- הרישום לקורס יתבצע באמצעות חברת ערן שחם בע"מ, עבור אוניברסיטת בר-אילן.



תעודות לדוגמא



לימודי תעודה
03-7384481
biu-es.ac.il

קמפוס חרדי
077-2753094/8
DesigProg.biu.ac.il

מכינה קדם אקדמית
03-5317956
mechina-kda.biu.ac.il

המדור לזרועות הביטחון
03-5317005/6
mzb.biu.ac.il

המדור לתוכניות מובנות
03-5317957
DesigProg.biu.ac.il

משרד ראשי
03-5318254
DesigProg.biu.ac.il

תוכנית 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 העיסקיים 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 UNION, INNER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN.
- שינוי טבלה ALTER, יצירה של אינדקסים, טבלאות זמניות.
- תכנות SQL מתקדם.
- סינון ומיון של נתונים ומשתנים 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 למדעני 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: (MS SQL Server, MySQL, PostgreSQL, Oracle, DB2,) רלציונליים
- (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): מנגנון 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 בייסיאני 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 היפר הפרמטרים, 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.
- בנייה ואימון של רשת פשוטה באמצעות KERAS.
- רשתות קונולוציונליות 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- TEXT TO IMAGE Diffusion Models 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 ביצירת 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!

