



Certified Data Scientist

לימודי תעודה להכשרת מדעני נתונים

עם התמחות ב-Machine Learning



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 מנתחי ומדעני נתונים עם התמחות

Machine Learning-ב

מנהל אקדמי: מר ערן שחם, המרכז למקצועות ההייטק באוניברסיטה
ראש התוכנית: ד"ר תומס קרפטי, מרצה בכיר וחוקר בינלאומי בתחום

שלחתם קורות חיים למשרה ומעולם לא קיבלתם תשובה. מה קרה שם בעצם?
ידעתם שבמחלקות משאבי אנוש נערך מחקר שלכם ואיסוף מידע על כל מועמד לתפקיד, כך שקורות החיים הם רק קצה הקרחון של המידע שניתן לקבל.
כיום מידע על מועמדים נאסף מפעילותם ברשתות חברתיות, בלוגים, תמונות ופעילות היסטורית באינטרנט. מכל המידע שנאסף נבנה פרופיל אישי ורק אז נבחנת ההתאמה למשרה.
את התפקיד הזה בחברה מבצע ה **Data Scientist** - תפקיד שהפך לאחד המאתגרים והמתגמלים בחברות בשנים האחרונות.
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 מסוגר לחזות את העתיד בהנתן 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.
 - בתוכנית נכיר את כל אלו ועוד. נכיר את הכלים איתם עובד ה **Data Scientist** בהתבסס על הכרות עם **Microsoft SQL Server** ושפות הפיתוח **Python** ו-**R**. בסיום הלימודים יגיש הסטודנט **עבודת גמר המתבססת על מחקר נתונים** והכוללת את כל שלבי הפיתוח וההטמעה שנלמדו בתוכנית.

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 התחילתו 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, ממוקדים **Microsoft SQL Server** ושפות הפיתוח **Python ו-R**.
- תעודה יוקרתית של אוניברסיטת בר אילן, המערך לתוכניות ייעודיות.
- קורס מעשי במהותו, המשלב יישום תרגולים רבים ובניית פרויקט אישי במקביל לידע התיאורטי הנרכש.
- הקורס מועבר בהדרכתם של מרצים מובילים וותיקים המאושרים ע"י אוניברסיטת בר אילן.
- סדנת מיתוג אישי לצורך מיקסום אפשרויות תעסוקה
- סביבות תרגול וירטואליות מתקדמות שהוקמו במיוחד לכל מודול.
- ליווי אישי של התלמיד לאורך כל המסלול וגם לאחריו.

תנאי קדם

- תואר ראשון או רקע אקדמי בתחומי מערכות מידע/תעשייה וניהול/מדעי המחשב/כלכלה/תחומי המדעים
- ראיון אישי ומבחן התאמה

מתכונת לימודים

- פתיחת הלימודים: **17.10.21**
- משך הלימודים: כ- 7 חודשים ויתקיימו בימים ראשון ורביעי בין השעות 17:30-21:30
- תכנית הלימודים כוללת **260 שעות אקדמיות**.

שכר לימוד

- דמי רישום: 400 ₪.
- שכר לימוד: 14,500 ₪.
- חניה: עבור תשלום נוסף של 200 ₪ ניתן לחנות את הרכב באוניברסיטה בשעות הלימוד עד סוף ספטמבר.

מקום הלימוד

- הלימודים נערכים במרכז למקצועות ההייטק, קמפוס אוניברסיטת בר-אילן, רמת גן

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

לפרטים והרשמה נא לפנות אל:

ערן שחם

050-6560666

eran.shaham1@biu.ac.il



תכנית הלימודים

מיתוג אישי בשוק התעסוקה המשתנה (5 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

מודול 1 - מבוא למדעי הנתונים (5 ש"א):

- מהו מדע הנתונים
- מחקר מדעי - שיטות מחקר
- אנטומיה של פרויקט במדעי הנתונים

מודול 2 - תכנות 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), לולאות ב-Python, פלט קלט (קריאת וכתיבת קבצים), עבודה עם סביבות וירטואליות- Virtual environment (יצירת סביבה וירטואלית, התקנת חבילות בסביבות וירטואליות, אופן עבודה בסביבה וירטואלית), SciKit ולמידת מכונה ML, ספריית NumPy (NumPy arrays), גרפים עם Matplot, Pandas ותזרימי שליטה- Data flow control (Frames), SciKit, Keras- TensorFlow.

מודול 3 - 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 ו-Docker containers.

מודול 4 - בסיס נתונים ותכנות SQL (15 ש"א):

מבוא ל-SQL (חיבור למסדי נתונים, יצירה ומחיקה של טבלאות, פקודות: SELECT, INSERT, DELETE, UPDATE), תכנות SQL מתקדם 1 (חיבור טבלאות, JOINS: INNER JOIN, LEFT OUTER JOIN, RIGHT OUTER JOIN, UNION. שינוי טבלה ALTER, יצירת אינדקסים, טבלאות זמניות), תכנות SQL מתקדם 2 (סינון ומיון נתונים, משתנים מחושבים: טרנספורמציות מתמטיות, פעולות על מחרוזת, פעולות על התאריך והשעה. אגרגציות של נתונים, פיבוטים PIVOT), פונקציות, Stored Procedures ותכנות דינמי.

מודול 5 - 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 SAS, SPSS, Json, Delimited, TAB, CSV
- 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): מנגנון לאחסון ואחזור נתונים, אחסון ערך-מפתח (Key-value), אחסון מסמכים, אחסון גרפים, אחסון אובייקטים.

מודול 6 - מבוא לסטטיסטיקה עם תכנות R (40 ש"א):

מבוא לתכנות R, סטודיו R, תחביר R, מניפולציה של משתנים (וקטורים, פקטורים, Data Frames, Data Tables), אופרטורים ב-R, לולאות, פונקציות, חבילות למדעי הנתונים, DPLYR ואגרציות, חקר נתונים באמצעות גרפים (ggplot2).

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

יצירת נתונים סינתטיים וסימולציות

מודלי רגרסיה: רגרסיה לינארית

מודלי קלסיפיקציה: רגרסיה לוגיסטית

רגרסיות תלויות זמן: ניתוח הישרדות

ניתוח סדרות עתיות: ARIMA

מודול 7 - למידת מכונה Machine Learning (135 ש"א):

הכנת נתונים:

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

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.

למידה עמוקה (Deep learning) - רשתות עצביות:

- רשתות עצביות ללא השגחה (Unsupervised): מכונת בולצמן מוגבלת Restricted Boltzman Machine, רשתות אמונה עמוקות Deep-belief networks, אוטו-אנקודרים Autoencoders.
- רשתות עצביות עם השגחה Supervised: רשתות Feedforward, Gradient descent, מעדכנים, פונקציות הפעלה activation functions, פונקציות הפסד loss function.
- רשתות חוזרות ו-LSTMs
- רשתות קונוולוציונליות convolutional networks

מודול 8 - הטמעה במערכת ייצור (15 ש"א):

- יישום במערך הייצור
- QA של יישום
- אימון והכשרה של לקוח

פרויקט מסכם (15 ש"א):

פרויקט הגמר- פרטוקול מחקר, תחקור נתונים וניקוי נתונים

- הנדסה ובחירת משתנים
- יצירת מודלי חיזוי והערכתם
- QA של תהליך מכונת למידה (ML)
- דו"ח מחקר
- יישום מודל החיזוי במערכת הייצור
- הדרכת משתמש הקצה להטמעה מיטבית