



TECHNION

Azrieli School of Continuing
Studies of the Technion

Big Data Analyst

ניהול אקדמי: מר אמיר בן חיים

מבוא

עם ההתפתחות הטכנולוגית בשנים האחרונות ועם הגידול בכמות המידע, השפעת התובנות העסקיות נהפכת למשמעותית יותר ויותר. ארגונים כיום שואפים להיות Data Driven Organization ולכן עולה הצורך באנשי מקצוע שידעו לארגן את המידע הנרחב והמגוון שנאסף מהערוצים השונים, להתאים אותו לצרכי הארגון, לעבד ולנתח אותו על מנת שיוכל לשמש את הארגון לקבל החלטות מושכלות המונעות ממידע. ניתוח נתוני עתק (Big Data) הוא תהליך מורכב של בחינה ועיבוד נתונים גדולים על מנת לחשוף תבניות והפיכת הדאטה הגולמי למידע חדש. כגון, דפוסים נסתרים, קורלציות (מתאמים) בין נתונים, מגמות שוק והעדפות של לקוחות. כל אלו הכרחיים לארגונים לקבל החלטות עסקיות נכונות. בשונה משיטות אחרות אשר עונות על שאלות בסיסיות אודות פעולות עסקיות וביצועים, ניתוח נתונים גדולים הוא תחקור מתקדם, הכולל יישומים מורכבים עם אלמנטים כגון מודלים חיזויים (Predictive Models), אלגוריתמים סטטיסטיים (Statistical Algorithms) וניתוחי What-If analysis.



החשיבות היא אדירה:

- הזדמנויות למקורות הכנסה חדשים
- שיווק יעיל יותר
- שירות לקוחות טוב יותר
- יעילות תפעולית משופרת
- יתרונות תחרותיים על פני יריבים

מטרת התכנית

תכנית Technion Certified Big Data Analyst נועדה להכשיר סטודנטים לנתח נפחים הולכים וגדלים של דאטה. בוגרי התוכנית מפתחים טכניקות לעבודה עם דאטה מובנה (Structured), מובנה למחצה (Semi Structured) ולא מובנה (Unstructured) - שלרוב לא מנוצלים על ידי מערכות אחרות ואנליטיקה קונבנציונאליות. על מנת לחבר את נושאי הלימוד לנעשה בתעשייה, ייחשף הסטודנט לשימוש במערכות ייעודיות לכמויות מידע עצומות וביצוע תהליכי אנליזה. כמו כן, יתרגל הסטודנט ניתוח מידע, שיטות חיפוש במערכות ועבודה מול מערכת קבצים.

תנאי קדם

- בוגרי מנהל עסקים, כלכלה, חשבונאות, רו"ח, מערכות מידע, סטטיסטיקה, מדעי החיים מהנדסים ומועמדים בעלי רקע רלוונטי העוסקים בדאטה ומעוניינים בשלוב תהליכים עסקיים עם כלים טכנולוגיים.
- ראיון קבלה ע"י יועץ אקדמי.
- קריאה והבנה ברמה טובה של אנגלית טכנית.

לייווי מקצועי

- אמיר בן חיים, מנהל אקדמי
- מר אמיר בן חיים הוא בעל ניסיון רב בתחום מערכות המידע, חקר וניתוח נתונים בחברות גדולות.
- יועץ לחברות עסקיות, מעניק שירותי אינטגרציה, פיתוח, והדרכה בתחומי, Big Data Analysis, ML, GenAI, BI.
- מרצה מוביל בביה"ס ללימודי המשך של הטכניון והמנהל האקדמי של תוכניות Big Data Analysis ו-Generative AI.



יתרונות התכנית

- הסמכה יוקרתית של הטכניון.
- קורס מקיף לתחקור נתונים.
- מיקוד בכלים הנפוצים בתחום, ממוקדים Microsoft SQL Server 2022.
- הכרות עם Big Data לסוגיו השונים
 - Hadoop and Ecosystem
 - MongoDB
 - Spark
 - Hive
- מודול Advanced SQL אשר מכין לראיונות עבודה ואתגרים בעולם האמיתי.
- מודול Python שפת התחקור ועבודה עם נתונים הפופולארית ביותר, לרבות:
 - NumPy
 - Pandas
 - Matplotlib
 - Seaborn
 - Scilcit-learn
- A/B Testing - נממש ב-PYTHON מבחן סטטיסטי נפוץ.
- Power BI – כלי מבית Microsoft לניתוח נתונים מורכב.
- Power Query / Power Pivot - כלים מובילים לניתוח ותחקור נתונים.
- Tableau – מודול ייחודי של אחת הפלטפורמות המובילות בעולם.
- GIT/GitHub - יצירת Repository וניהול גרסאות
- מודול ChatGPT כלי AI בעל חשיבות אדירה
 - Code Interpreter
 - Plugins
 - Prompt Engineering
 - OpenAI API
- קורס מעשי, המשלב יישום תרגולים רבים ובניית פרויקט אישי לתיק העבודות במקביל לידע התיאורטי.
- מועבר בהדרכתם של מרצים מובילים וותיקים המאשרים ע"י מוסד הטכניון.
- סביבות תרגול וירטואליות מתקדמות שהוקמו במיוחד לכל מודל.
- ליווי אישי של התלמיד לאורך כל המסלול.



פירוט תכנית הלימודים

Introduction to Data Analysis + SQL Basics (15 hours)

- Familiarity with Data Analysis Field
- Functions and job role of Data Analyst
- Basic select statements
- Filtering and sorting the results
- Scalar functions
- Group functions
- Join methods
- Sub Queries
- Add, modify, or remove data from a table (DML)

Managing SQL Objects (10 hours)

- Creating and managing tables (DDL)
- Creating and using views
- Working with Indexes
- Security - Windows Authentication & Sql Server Authentication

T-SQL Programming (10 hours)

- Working with scalar variables and table variables
- Loops and conditions
- Creating and using stored procedures
- Creating and using functions (User-Defined Functions)

Advanced SQL Querying (10 hours)

- Using CTEs
- Recursion
- Advanced usage of subqueries
- Window Function
- Analytic functions – Lag, Lead, First_Value
- Rank Functions - Rank, Row_Number, Dense_Rank
- Pivot and Unpivot
- Union / Union All / Intersect / Except

Familiarity with BI, Business intelligence fundamentals (5 hours)

- Introduction to DWH concepts
- Designing a Dimensional Model - Normalization vs Star & Snowflake Schemas
- Designing the ETL process
- Introduction to Tabular Model and best practices

Power BI- Advanced Dax (15 hours)

- SSAS Tabular & POWER BI similarities and differences
- Working with Measures, Calculated Columns, Calculated Tables, Row Filters
- DAX Formulas and Dax Queries
 - Simple Aggregative functions
 - Harnessing model relationships with "Related" & "RelatedTable"
 - The "X" functions-row iterators
 - Using Filter, Values & All
 - Time Intelligent functions
 - Calculate & Calculate Table
- Managing data and creating reports in Microsoft Power BI Desktop
- Power BI - Import / Direct Query modes
- Reporting best practices: choosing the right layout & visuals
- Slicers, Filters, and visual interaction
- Using Buttons & Bookmarks
- Power BI special features:
 - "what-If" ◦ "External visuals from the market"
 - "Explain the decrease & increase" ◦ Custom Tooltip"
 - "Quick measures" ◦ "External visuals from the market"
 - Custom Tooltip"
- Power BI Dashboards
- Understanding the differences between Power BI Service & Power BI Report Server
- Publishing Power BI to the cloud
- Dax Studio

Power Pivot / Power Query - Microsoft data analysis tools (10 hours)

- Connecting to data
- Extract Transform Load
- Data mashup (get & transform) multiple sources using Power Query ("M")
 - Get Date from files, DB & folders
 - Filtering, merging & manipulating Data using vast transformation wizards
 - Using "M" language to edit queries with advanced editor
 - Using parameters & functions
- Creating a Model
- Creating relations between tables
- Creating measures and Kpi's
- Working with Power Pivot Table and Power Pivot Charts

Tableau (15 hours)

- Tableau Prep
 - Creating Connection to Data Source
 - Organizing Your Data
 - Manipulate Data
- Desktop Basics
 - Tabs
 - Status bar
 - Panes
 - Cards
 - Toolbar/Menus
 - Show Me
 - Full Screen
- Creating Report & Dashboards
 - Understanding Fields
 - Basic Worksheet Options
 - Filtering & Sorting
 - Organizing Data
 - Displaying your data
 - Understanding Fields
 - Viewing Data
 - Parameters
 - Chart types
- Tableau Formulas
 - Creating and Editing Calculated Fields
 - Aggregations, String Functions, Date Calculations
 - Table Calculations
- Defining subsets of your data

Statistics for Data Analysis - Basics (5 hours)

- Combinatorics
- Probability
- Conditional probability & Bayesian
- A/B testing

Python Basics (10 hours)

- The working environment
- Basic data types
 - Numbers
 - Strings
 - Booleans
 - None
- Collections
 - Lists
 - Tuples
 - Dictionaries
 - Sets
 - Frozenset
- Control flow
 - if...else
 - for...in
 - list comprehension
 - while
- Functions
 - def fun
 - lambda
- Classes
- Modules

Python NumPy Library (5 hours)

- Basic concepts
- Working with arrays

Python Matplotlib & Seaborn Library (10 hours)

- Basic concepts
- Creating a basic plot
- Other common plots
- Advanced plot formatting

Python Pandas Library (10 hours)

- Basic concepts
- Series and Index
- Data Frame
- Aggregations and advanced manipulations
- Group By
- Visualizations

Python EDA (Exploratory data analysis) (10 hours)

- Hands on and use cases

Python Connecting to Data Resources (5 hours)

- Intro to working with DBs with SQLAlchemy package
- Getting data from files and DB

Big Data Solution Introduction (30 hours)

- Introduction to Big Data
- Big Data Architecture and NoSQL technologies
- Technologies and trends in the world of Big Data
- Hadoop Infrastructure and Ecosystem Services
- Hadoop core and ecosystem - The essentials (e.g. HDFS)
 - Data ingestion technologies – Overview (Sqoop, Flume)
 - Using Hive to query HDFS data with SQL
- Spark
 - Introduction to Spark and PySpark
 - The Basics
 - Spark Sql
 - Spark Streaming
- Unstructured data - Introduction to MongoDB
 - Core concepts
 - Environments
 - CRUD and the MongoDB Shell
 - Understand documents and collections
 - Querying MongoDB database
 - Understanding MongoDB replication and sharding

GIT/GitHub (5 hours)

- Create local repository
- Git Actions
- Connect to GITHUB with GIT
- Introduction to Version Control
- Using Git Locally
- Working with Remotes
- Collaboration

Introduction to Data Science & Machine Learning (10 hours)

- Data Science overview
- Technologies and trends
- Machine learning Architecture and core concepts
- ML algorithms - Decision Tree, Random Forest, Knn
- Intro to NLP
- Python Implementation for ML

ChatGPT (10 hours)

- Introduction
 - Generative AI & LLM
 - OpenAI and ChatGPT
 - Terminology
- ChatGPT Prompt Engineering
 - Prompts: clear and specific
 - Context and instructions
 - Prompt formats: completion, classification, question & answering
 - Fine-tuning prompt parameters
- Advanced Data Analysis (Python & Sql)
 - Create
 - Improve
 - Optimise
 - Explain
 - Debugging
 - Execute
 - Analyse
- Introduction to OpenAI API
 - Overview of OpenAI API
 - Using the OpenAI API (using Python)
 - Registration, authentication, API keys, API calls
 - Role's: system, user, assistant, function
 - API Parameters: temperature, max tokens, top p, top k values
 - Tune API parameters
 - Intro to fine-tuning customize model
- ChatGPT plugins
 - Integration
 - Data processing
 - Extended capabilities

מועד פתיחת התכנית / שעות לימוד

פתיחת הלימודים: 09.07.2024 יום שלישי

ימי שלישי בין השעות 17:30-21:30 וימי שישי בין השעות 09:00-13:00.

מתכונת הלימודים

תכנית הלימודים מורכבת מ-200 שעות אקדמיות (40 מפגשים), לימודים פרונטאליים בקמפוס שרונה בת"א (22 מפגשים) בשילוב מפגשים דיגיטליים (18 מפגשים).

שיעורי הלימוד הדיגיטליים (הקלטות) פותחו במיוחד ללמידה עצמאית מהבית.

השיעורים הדיגיטליים יהיו זמינים החל ממועד השיעור.

חומרי הקורס לרבות ההקלטות יהיו זמינים לשימוש למשך הקורס ועד כחודש לאחריו (הקלטות, מצגות, מטלות-זמינים לשימוש ללא הורדה).

שאלות / תמיכה לסטודנטים ייענו במפגשים הפרונטאליים.

התכנית נלמדת במעבדת מחשבים עם מחשבי PC ובמערכת הפעלה windows.

הן המפגשים הפרונטאליים והדיגיטליים, כוללים הפסקות וזמן תרגול עצמאי.

זכאות לתעודה

תוענק תעודת גמר של בית הספר ללימודי המשך של הטכניון.

סיום הלימודים וקבלת תעודת גמר מותנה:

- בנוכחות ב-80% מהמפגשים לפחות
- עמידה בכל מטלות החובה בקורס לרבות פרויקטים, מטלות, מבחנים
- הגשת תרגילים און-ליין בשיעורי הלימוד הדיגיטליים

עלות התכנית

דמי רישום: 480 ₪

שכר לימוד: 14,200 ₪

מקום הלימוד

קמפוס הטכניון בת"א, מתחם שרונה, רח' דוד אלעזר (מול הקריה וליד תחנת רכבת השלום).

לפרטים נוספים והרשמה:

מעין זהבי | רכזת תכניות המחשבים

maayanz@trdf.technion.ac.il | 077-8871701

הערות

פתיחת כל תכנית מותנית במספר הנרשמים | דמי ההרשמה אינם מוחזרים, אלא במקרה של אי פתיחת התכנית על ידי בית הספר | דמי ההרשמה אינם כלולים בשכר הלימוד | בית הספר מביא לידיעת הנרשמים, כי ייתכנו שינויים במערך התכנית, במועדי הלימודים והבחינות או בכל נושא אחר. הודעה על כל שינוי תימסר למשתתפים | כפוף לנהלי בית הספר ללימודי המשך במוסד הטכניון למו"פ בע"מ.