

## POWER ACADEMY

### Machine Learning for Developers, BI Professionals and Business Analysts

#### אודות הקורס

עולם הידע ב- Machine Learning עוסק באלגוריתמים ושיטות לבניית מערכות לומדות, שיכולות להשתמש במידע שנאסף על ידי הארגון, על מנת להפיק באופן אוטומטי ידע ותובנות חדשים.

שליטה ב- Machine Learning, מייצרת לארגונים הזדמנויות חדשות ומרגשות להשיג יתרון תחרותי, על ידי תכנון אלגוריתמים שיסייעו בקבלת החלטות ארגונית טובה יותר. ב- Machine Learning, אופן השימוש במידע הארגוני הוא עמוק יותר, שיטתי יותר ומתבסס על מניפולציות מתמטיות מורכבות.

הקורס כולל שני חלקים (מודולים):

חלק ראשון (נדרש לכל המשתתפים בקורס): היכרות עם מתודולוגית העבודה ב- Machine Learning, ועם השיטות והאלגוריתמים המרכזיים המשמשים כחלק ממתודולוגית זו. חלק זה ילווה בתרגול ובדוגמאות רבות, תוך שימוש ב- R, שפת Script נפוצה ושימושית ל- Machine Learning.

חלק שני (מתאים למפתחים ואנשי BI שיעסקו בהטמעת פתרונות Machine Learning כחלק ממערכות Production): בחלק זה ניקח בחשבון את מגוון השיקולים הנדרשים במעבר מאלגוריתם קונספטואלי לפתרון בעיה אל מערכת גדולה, הנדרשת לפעול על כמויות נתונים גדולות, באופן אוטומטי וביעילות.

#### מטרות הקורס

מטרת הקורס הנה לחשוף את הסטודנטים לעולם המתפתח והמרגש של *Machine Learning*, תוך שימת דגש על הפן העסקי והפרקטי, על מנת לאפשר לבוגרים ליישם את השיטות הנלמדות בארגוניהם, הלכה למעשה.

## קהל יעד

- מפתחי תוכנה
- מפתחי Big Data
- אנשי BI
- אנליסטים עסקיים

## דרישות קדם

עבור חלק ראשון-

- הבנה סטטיסטית ומתמטית בסיסית (ברמה של קורס מבוא לסטטיסטיקה)
- ניסיון בכתיבת קוד בשפה כלשהי, אפשר גם במסגרת לימודים (הקורס מתאים גם למי שאינו מפתח תוכנה בעל ניסיון)

עבור חלק שני-

- ניסיון מעשי בכתיבת תוכנה בשפת object-oriented כלשהי (רצוי ב-Java) או ניסיון ב-scala
- היכרות טובה עם בסיסי נתונים טבלאיים.

## משך הקורס

40 שעות אקדמאיות

## **תכני הקורס**

### **חלק ראשון (32 שעות)**

- הקדמה
  - מושגים ראשוניים
  - מוטיבציה עסקית
  - סקירת סוגי בעיות ב-Machine Learning
  - דוגמאות לשימוש ב-Machine Learning בארגונים.
- מתודולוגיית CRISP-DM
  - מחזור העבודה האופייני בפרויקטי Machine Learning
  - התמקדות בהטמעת פתרונות Machine Learning בסביבת הפיתוח.

Initial Data Understanding ▪

Correlations ○

Outliers Detection ○

Data Types ○

Data Preparation ▪

Data Normalization ○

Feature Selection ○

Feature Extraction ○

Supervised Learning ▪

Generalization ○

Bias ○

Variance ○

Classification ○

Regression ○

Basic Supervised Learning Algorithms ▪

K-NN ○

Decision Trees ○

Linear Regression ○

Logistic Regression ○

SVM ○

Naïve Bayes ○

Basic Unsupervised Learning ▪

K-Means ○

EM ○

Advanced Methods ▪

Ensembles ○

Neural Networks ○

Deep Learning ○

- Model evaluation ▪
  - train-test-validation ○
  - cross validation ○
  - Confusion Matrix over\under-fitting ○
  - Accuracy ○
  - ROC Curve ○
  - Lift Charts ○

### חלק שני (8 שעות)

- Architecture considerations for building ML pipelines at scale ▪
  - Challenges ○
  - Lambda architecture ○
  - Choosing the right db source ○

- Spark as distributed computing ML framework ▪
  - Spark intro ○
  - Spark MLlib overview and practice ○

- Common ML cloud services ▪
  - Amazon machine learning service ○
  - Google prediction api ○