The background features a large diagonal split, with the top-left portion being black and the bottom-right portion being light blue. Scattered across the entire background are various hexagonal shapes. Some are solid black, some are solid light blue, and many are white with black or blue outlines. Some of these hexagons are interconnected by thin lines, creating a network-like pattern. The overall aesthetic is modern and technical.

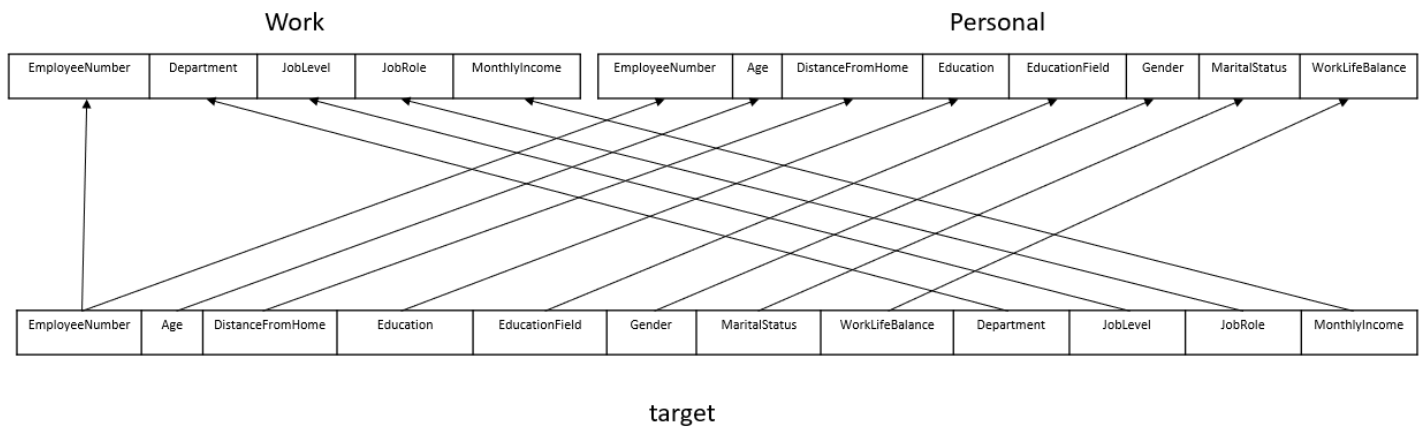
בינה עסקית – פרויקט חלק 3

מגישות
ורוניקה פרידמן
לי קישון

סמסטר ב, 2022

חלק א': STTM

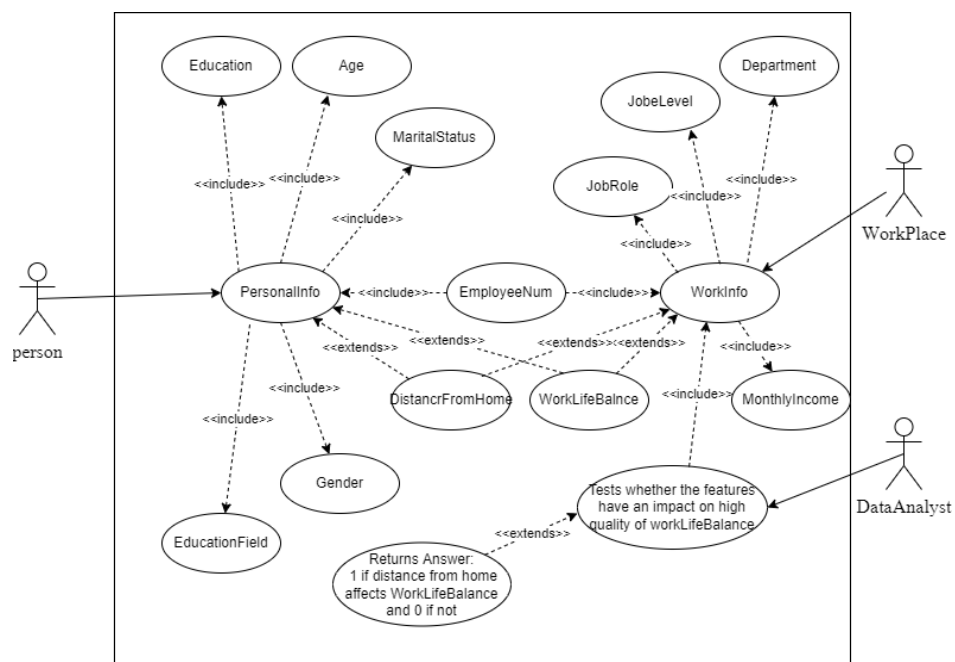
1. סכימה ויזואלית אשר מציגה את ה - STTM :



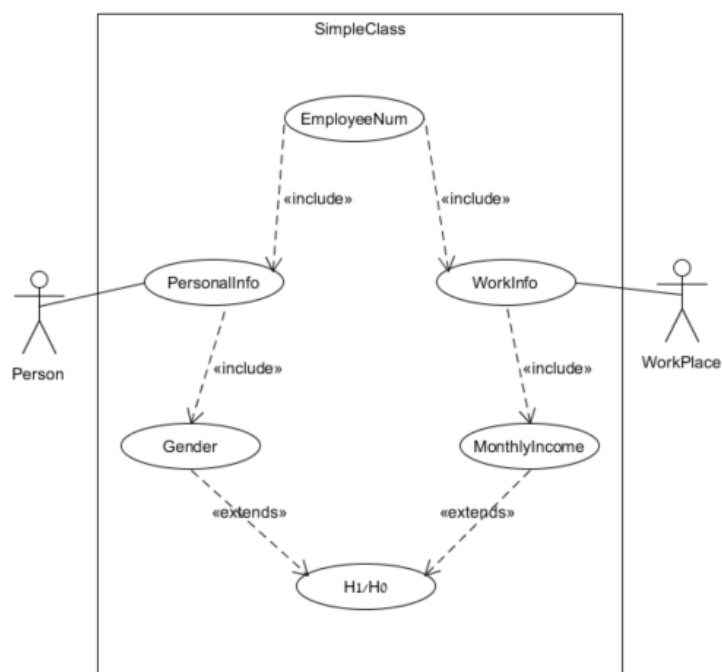
2. מצ"ב בקובץ אקסל.

חלק ב': data mining techniques

1. תהליך ה - KDD אשר אנו מבצעים בפרויקט זה הוא בשיטה ה - Descriptive מאחר ורוב ה - Data מחולק לקבוצות, וחסר מידע עסקי קודם. ולכן על מנת לענות על השאלות העסקיות שלנו אנחנו נדרשים לנתח את הנתונים והמידע הקיימים לרשותינו.
2. הטכניקות אשר ימומשו על אוסף הנתונים הם :
 - clustering, סוג הנתונים הם ברובם מספרים נומריים וגם בינארים הנדרשים לחלוקה לקבוצות.
 - 2 תרחישי USE CASE :
 - שאלה עסקית 1:



■ שאלה עסקית 2:



3. הגדרת מדד דמיון עבור הDW שבנינו:

○ Jaccard - עבור הנתונים הקטגוריים לרבות בינאריים כאשר הממד יוגדר בין 0 ל-1 (1 דמיון מלא, 0- אין דמיון).

○ Sorensen-Dice - עבור משתנים בדידים.

4. השערות עבור כל אחת מהשאלות העסקיות:

שאלה עסקית: האם התכונות הבאות: גיל, מרחק מהבית, השכלה, מגדר ומצב משפחתי - בעלות השפעה על איכות חיים-עבודה גבוה (מעל 2)

○ H0- Distance from home affects Worklifebalance rate

○ H1- Distance from home does not affects Worklifebalance rate

○ דרך קבלת ההחלטה: Sorensen-Dice

○ מדוע זו הדרך: מרחק מהבית הינו נתון בדיד ולכן השתמשנו בדרך זו.

שאלה עסקית: מה ניתן ללמוד על תכונות נתוני ההעסקה - (Department, JobLevel, JobRole, MonthlyIncome) של העובדים ביחס למגדר שלהם.

○ H0 - men and women get the same MonthlyIncome

○ H1- Men's MonthlyIncome is higher than a Women's MonthlyIncome

○ דרך קבלת ההחלטה: מבחן חי בריבוע

○ מדוע זו הדרך: מאחר והמדד העיקרי הינו קטגוריאלי מבחן חי בריבוע הוא המתאים ביותר כאשר ממוצע שכר הגברים הינו הערך המצופה וממוצע שכר הנשים הינו הערך הנצפה.

חלק ג': שאלות

מצורף בקובץ אקסל בתיקיית PIPELINE.

SELECT AVG (MonthlyIncome) as avg_MonthlyIncome ,Gender FROM WorkData INNER JOIN PersonalData ON EmployeeNumber=EmployeeNumber
SELECT COUNT(WorkLifeBalance) as Rate FROM WorkData WHERE WorkData=Gender
SELECT AVG (JobLevel) as avg_JobLevel ,Gender FROM WorkData INNER JOIN PersonalData ON EmployeeNumber=EmployeeNumber
SELECT MAX(Age) as max_Age, MIN(Age) as min_Age, AVG (MonthlyIncome) as avg_Age FROM WorkLifeBalance WHERE WorkLifeBalance>1
SELECT MAX(DistanceFromHome) as max_DistanceFromHome, MIN(DistanceFromHome) as min_DistanceFromHome, AVG (DistanceFromHome) as avg_DistanceFromHome FROM WorkLifeBalance WHERE WorkLifeBalance>1
SELECT MAX(Education) as max_Education, MIN(Education) as min_Education, AVG (Education) as avg_Education FROM WorkLifeBalance WHERE WorkLifeBalance>1

חלק ד': ניהול גרסאות

1. קישור לפרויקט:

<https://github.com/VeronikaFridman/BI-project-HR-IBM.git>