תרגיל 3 - Python

1. טען את קובץ נתוני האימון train_Loan.csv
2. הדפס את התפלגות הערכים (Frequency Distribution) של המשתנים השונים בקובץ האימון.
3. הדפס את הטיפוס (Type) של כל אחד מהשתנים בקובץ האימון.
4. בצע תיקון Missing Values עבור המשתנים Gender, Married, ו- Self_Employed.
5. בצע דיסקרטיזציה (מכל סוג) למשתנה LoanAmount. ציין את ה-bins השונים.
6. בצע Outlier Detection למשתנה LoanAmount, על פי 3 סטיות תקן בהתפלגות המשתנה.
7. הוסף עמודה חדשה בשם Normalized_Income לקובץ אימון אשר מיוצגת ע"י הנוסחה: 0.5*sqrt('Applicant_Income').
8. צור Dummy Variable עבור המשתנה Education, עבורו הערך 'Graduate' יקבל את הקידוד '1', והערך 'Not Graduate' יקבל את הקידוד '0' .
9. הוסף את תוצאות משתנה ה- Dummy Variable ל- Dataset המכיל את נתוני הטבלה המקורית.
10. הפוך משתנים קטגוריים למשתנים נומרים על מנת לאפשר שימוש בספרייה SK-Learn.
11. הפעל את ה- Decision Tree Classifier על קובץ האימון, ואמן את המודל תוך שימוש ב- 10-fold Cross Validation.
12. ציין את דיוק האימון של המודל (Accuracy) ואת ה- Cross Validation Score
13. צייר את עץ ההחלטה באמצעות הספרייה GraphViz
14. ייצא את טבלת נתוני האימון המעודכנת לקובץ בשם train_Loan_updated.csv

הוראות הגשה

א. אי עמידה בכל אחת מההוראות יגרור הורדת ציון או פסילת העבודה.
ב. הגשת העבודה ב**זוגות** בלבד. רק אחד מבני הזוג יגיש את המטלה!
ג. שפת תכנות – Python 2.7 ומעלה, סביבת פיתוח – מומלץ להשתמש ב-JetBrains PyCharm גרסה 2019.1 ומעלה.
ד. יש להגיש את העבודה **לתיקיית ההגשה הרלוונטית באתר הקורס (Moodle).**
באחריותכם האישית לבדוק לפני הגשה כי כל הקבצים נפתחים כראוי.
ה. **יש להגיש קובץ ZIP** - שם הקובץ יהיה מורכב משני מספרי תעודות הזהות של המגישים באופן הבא: ID1_ID2.zip
הקובץ יכיל את הקבצים הבאים:
- קבצי קוד Python עבור כל אחד מהסעיפים 1-14.
- קובץ word המכיל צילום מסך של תוצאות ההרצה ב- Python של הסעיפים 11-13.
- קובץ בשם train_Loan_updated.csv המכיל את טבלת נתוני האימון המעודכנת.

ו. אין לשתף קטעי קוד ואין להעתיק פתרונות!
ז. בנוסף, זוהי עבודה תכנותית ולפיכך יהיה משקל לכך בבדיקה. כלומר: יש לדאוג לקוד מסודר, הערות בקוד, לשמות משתנים בעלי משמעות וכדומה. יש לחלק את הקוד לפונקציות (במידת האפשר ולפי הצורך).
ח. שאלות בנוגע לתרגיל יש לשאול **אך ורק** בפורום השאלות הרלוונטי המופיע ב-moodle (ולא במייל - שאלות במייל לא יענו).

בהצלחה !!