**עבודת הגשה #4 בינה עסקית ומדעי הנתונים:**

**מבנה הפרויקט:**

בחלון הנפתח בעת הרצת התוכנית יש לבחור את הנתיב המלא אל התיקייה בה נמצאים קבצי ה- train, ה- test וה- Structure. לאחר מכן יש לכתוב את גודל ה- bind הרצוי ואז תתאפשר לחיצה על כפתור ה- Build. ברגע שבניית המודל נגמרת תעלה התראה שהבנייה הסתיימה ולאחר מכן יהיה ניתן ללחוץ על כפתור ה- classify ולקבל את פלט הסיווג.

**תיאור המחלקות:**

* המחלקה NaiveBayes: מחלקה זו הינה ה"מוח" של התוכנית, בה מתבצעים שלבי ה- build וה- classify.
* המחלקה GUI: כשמה כן היא מכילה את קוד ה- GUI של התוכנית.

**תיאור השיטות:**

* Build (folderPath, bind): פונקציה זו קוראת את קובץ ה- train וה- structure ובונה מהם את מבני הנתונים בהם אנו משתמשים בתוכנית. לאחר מכן מתבצעת השלמת ערכים חסרים בכל רשומה (בעזרת הפונקציות fillEmptyValues ו- (missingValues , ולבסוף מתבצעת דיסקרטיזציה לערכי התכונות הנומריות (בעזרת הפונקציה discretization).
* Classify(): פונקציה בה מתבצע תהליך הסיווג ע"י חישוב הסתברויות. בתחילת הפונקציה נקרא קובץ ה- test ובדומה לפונקציה הקודמת מתבצעת השלמת ערכים חסרים ודיסקרטיזציה. לאחר מכן נעשה חישוב ההסתברויות בעבור כל רשומה בקובץ ה- test לפי נוסחת ה- M-Estimator (המתבצעת בפונקציה mEstimator והיא נעזרת בפונקציה ncValue) ומודפס את קובץ ה- output הסיווג הנבחר של אותה רשומה.