

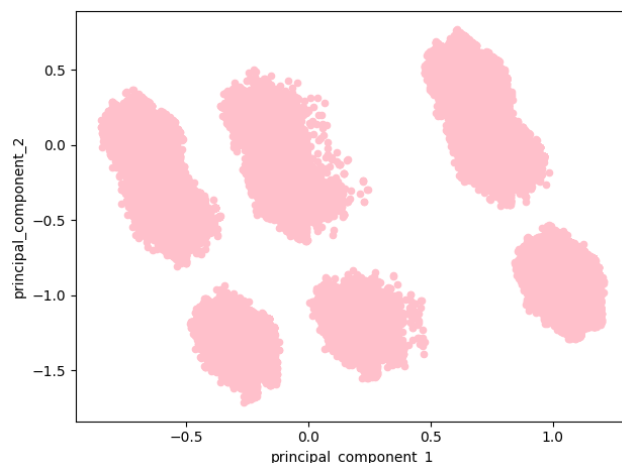
ניהול מידע מבוזר - פרויקט חלק ב'

Static Data Analysis – Feature Extraction

- אנחנו מנרמלים את הערכיים הנומריים על מנת להבטיח שכל העמודות הנומריות יהיו באותו סדר גודל, כך שאף פיצ'ר לא יקבל משקל יתר רק בגלל שהוא נמדד ביחידות גדולות יותר או בטווח ערכים רחב יותר. זה מקל על השוואה בין משתנים, משפר יציבות חישובית, ומאפשר לאלגוריתמים להתייחס לכל הפיצ'רים בצורה הוגנת ואחידה.
- כאשר מקצים ערכים מספריים ישירים לקטגוריות, האלגוריתם עלול לפרש את המספרים כאילו יש ביניהם סדר או מרחק בעלי משמעות, אף שבפועל מדובר בנתונים נומינליים בלבד. מצב זה עלול לגרום להטיה באופן שבו המידע הקטגורי מיוצג ומשפיע על תוצאות הניתוח. באמצעות שיטת One-hot encoding, כל הקטגוריות מקבלות ייצוג עצמאי ואחיד, ללא סדר מדומה או קשרים כמותיים שאינם קיימים במציאות, ובכך נשמרת המשמעות המקורית של הנתונים הקטגוריים.

Static Data Analysis – Visual Analysis

- ניתן לראות בתרשים 6 קבוצות של נקודות ורודות, כאשר יש 2 שלשות של קבוצות "דומות" בצורתן, ולכל קבוצה יש צפיפות שונה. כלומר $c=6$.



Static Data Analysis – Cluster's Viewing Analysis

- באופן כללי, ערך הdiff_rank מצביע על ההבדלים בדירוגים של התחנה בתת הקבוצה (הקלסטור), בהשוואה לקבוצה הכללית. אם ערך זה הוא חיובי, משמע בתת הקבוצה התחנה מדורגת גבוה יותר ביחס לקבוצה הכללית. אם ערך זה הוא שלילי, משמע בתת הקבוצה התחנה מדורגת נמוך יותר ביחס לקבוצה הכללית.
- השוואת תתי הקבוצות בין הקלסטורים השונים מצביעה על הבדלים מובהקים בזהות התחנות המובילות וברמות הפערים ביניהן, מה שמעיד על שונות בהעדפות התוכן בין הבתים. בנוסף, ההשוואה בין תתי הקבוצות בתוך אותו אשכול מראה כי הקבוצה שמכילה את כל משקי הבית השלישיים אכן מדגם יחסית מייצג אל מול הדאטה השלם. ניתן לראות מהטבלאות כי הקלאסטרינג משקף בפועל הבדלים בתוכן המועדף על קבוצות הצופים, שכן בכל קלאסטר מופיעות העדפות אחרות.

Dynamic Data Analysis – Streaming

- השוואת הקלסטורים בין הגלים – לעיתים ניתן לראות עבור אשכולות מסוימים כי הערוצים הנצפים ביותר מופיעים בכל הגלים, לעיתים בסדר שונה, ולעיתים ממש מופיעות אותן תחנות מובילות עבור קלאסטר מסוים.
- ניתן לראות כי קיימים דפוסים ועקביות בין הערוצים שנבחרים בראש הטבלה.
- ניתן לראות כי הוספה של נתונים אכן משמרת לעיתים את ההעדפות, ולכן אפשר להגיד שיש העדפה מובהקת של משקי בית לתחנות מסוימות – כלומר זה הקשר בין משקי הבית לקלסטורים אליהם משויכים.

Competitive Bonus

- תוספת היוגורט האהובה על אניס – הכל !!!!
- הצבע האהוב על אביב – סגול
- ליהי – הפרוזן יוגורט המועדף עליי, הוא על בסיס אסאי / חלב פרה, עם תוספות של פירות וחלווה מלמעלה.
- נוי – מאז שאני זוכרת את עצמי הצבע האהוב עליי הוא ירוק. ירוק מסמל התחדשות, צמיחה, ורעננות.