

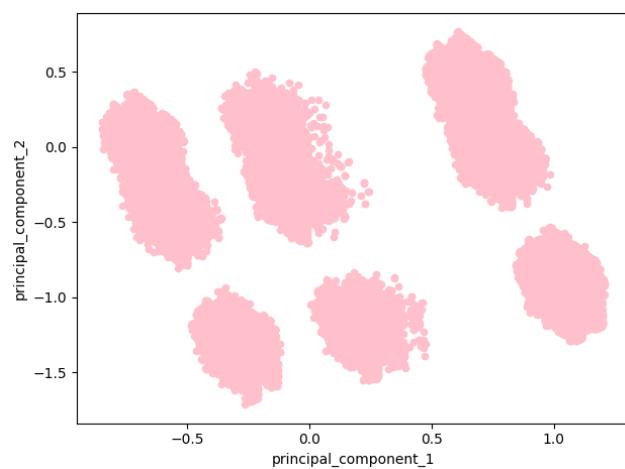
ניהול מידע מבוצר - פרויקט חלק ב'

Static Data Analysis – Feature Extraction

- אנחנו מנורמלים את הערכים 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 בטוחה ערכאים רחב יותר. זה מקל על השוואה בין משתנים, משפר יציבות חישובית, ומאפשר לאלגוריתמים להתייחס לכל הפיצ'רים בצורה הוגנת ו אחידה.
- כאשר מקצים ערכים מספריים ישירים לקטגוריות, האלגוריתם עלול לפרש את המספרים כאלה יש ביניהם סדר או מרחק בעלי משמעות, אף שבפועל מדובר בנתונים נומינליים בלבד. מצב זה עלול לגרום להטיה באופן שבו המידע הקטורי מיוצג ומשפיע על תוצאות הניתוח. באמצעות שיטת One-hot encoding, כל הקטגוריות מקבלות ייצוג עצמאי ו אחיד, ללא סדר מודומה או קשרים כמותיים שאינם קיימים במציאות, ובכך נשמרת המשמעות המקורית של הנתונים הקטוריים.

Static Data Analysis – Visual Analysis

- ניתן לראות בתתרשים 6 קבוצות של נקודות ורודות, כאשר יש 2 שלשות של קבוצות "זומות" בצורתן, ולכל קבוצה יש צפיפות שונה. קלומר $c=6$.



Static Data Analysis – Cluster's Viewing Analysis

- באופן כללי, ערך `rank_diff` מצביע על הבדלים בדירוגים של התנהה בתה הקבוצה (הקלסטר), בהשוואה לקבוצה הכללית. אם ערך זה הוא חיובי, משמע בתה הקבוצה התנהה מדורגת גבוהה יותר מאשר לקבוצה הכללית. אם ערך זה הוא שלילי, משמע בתה הקבוצה התנהה מדורגת נמוך יותר מאשר לקבוצה הכללית.
- השוואת תת-הקבוצות בין הקלסטרים השונים מצביעה על הבדלים מובהקים בזיהות התנהות המובילות וברמות הפערים ביניהן, מה שמעיד על שונות בעדפות התוכן בין הבטים. בנוסף, ההשוואה בין תת-הקבוצות בתווך אותו אשכול מראה כי הקבוצה שמכילה את כל משקי הבית השלישיים אכן מוגם וחסית מייצג אל מול הדעתה השלה. ניתן לראות מהטבלאות כי הקלאסטרינג משקף בפועל הבדלים בתוכן המועדף על קבוצות הצופים, שכן בכל קלאסטר מופיעות העדפות אחרות.

Dynamic Data Analysis – Streaming

- השוואת הקלסטרים בין הגלים – לעיתים ניתן לראות עבור אשכולות מסוימים כי הערכתם הנכפים ביוטר מופיעים בכל הגלים, לעיתים בסדר שונה, ולעיתים ממש מופיעות אותן תנהות מובילות עבור קלאסטר מסוים.
- ניתן לראות כי קיימים דפוסים ועקבות בין הערכתם שנבחרים בראש הטבלה.
- ניתן לראות כי הוספה של נתונים אכן לשמור לעיתים את העדפות, ולכן אפשר להגיד שיש העדפה מובהקת של משקי בית לתנהות מסוימות – ככלומר זה הקשר בין משקי הבית לקלסטרים אליהם משוויכים.

Competitive Bonus

- תוספת היוגרט האהובה על אניס – הכל !!!
- הצבע האהוב על אביב – סגול
- ליהי – הפרוזן יוגרט המועדף עליי, הוא על בסיס אסאי / חלב פרה, עם תוספות של פירות וחלואה מלמעלה.
- נוי – מאז אני זכרת את עצמי הצבע האהוב עליי הוא י록. י록 מסמל התחדשות, צמיחה, ורענן.