

פרויקט בלמידת מכונה וסטטיסטיקה ביישומים רפואיים

אפיון פנוטיפים קליניים של חולי קורונה בأחוזות למדיה לא מפוקחת

מגושים: עמית פילר ויוגב חדד

האם ניתן לזיהות פרופילים של סיכון בבדיקות דם שגרתיות?

2

השאלה שלנו

האם בבדיקות דם שגרתיות מכילות מידע מיידע חיוני המאפשר לזיהות קבוצות חולמים עם פרופילים ביולוגיים שונים על בסיס דפוסי בדיקות הדם בלבד?

1

הבעיה

האתגר הקליני אינו אבחן המחלה (חיובי/שלילי), אלא ניהול **הסיכון**. כיום, לרופא במילון חסרים כלים אובייקטיביים שיזהו מי מהחולים יחלים בקהלות ומיהולו לקרים תוך שעות

3

ההשערה

בדיקות דם שגרתיות המכילות מידע מיידע המאפשר הבחנה בין קבוצות חולמים עם דפוסים ביולוגיים שונים.

סינון רעים, טיפול בערכים חסרים ונרמול

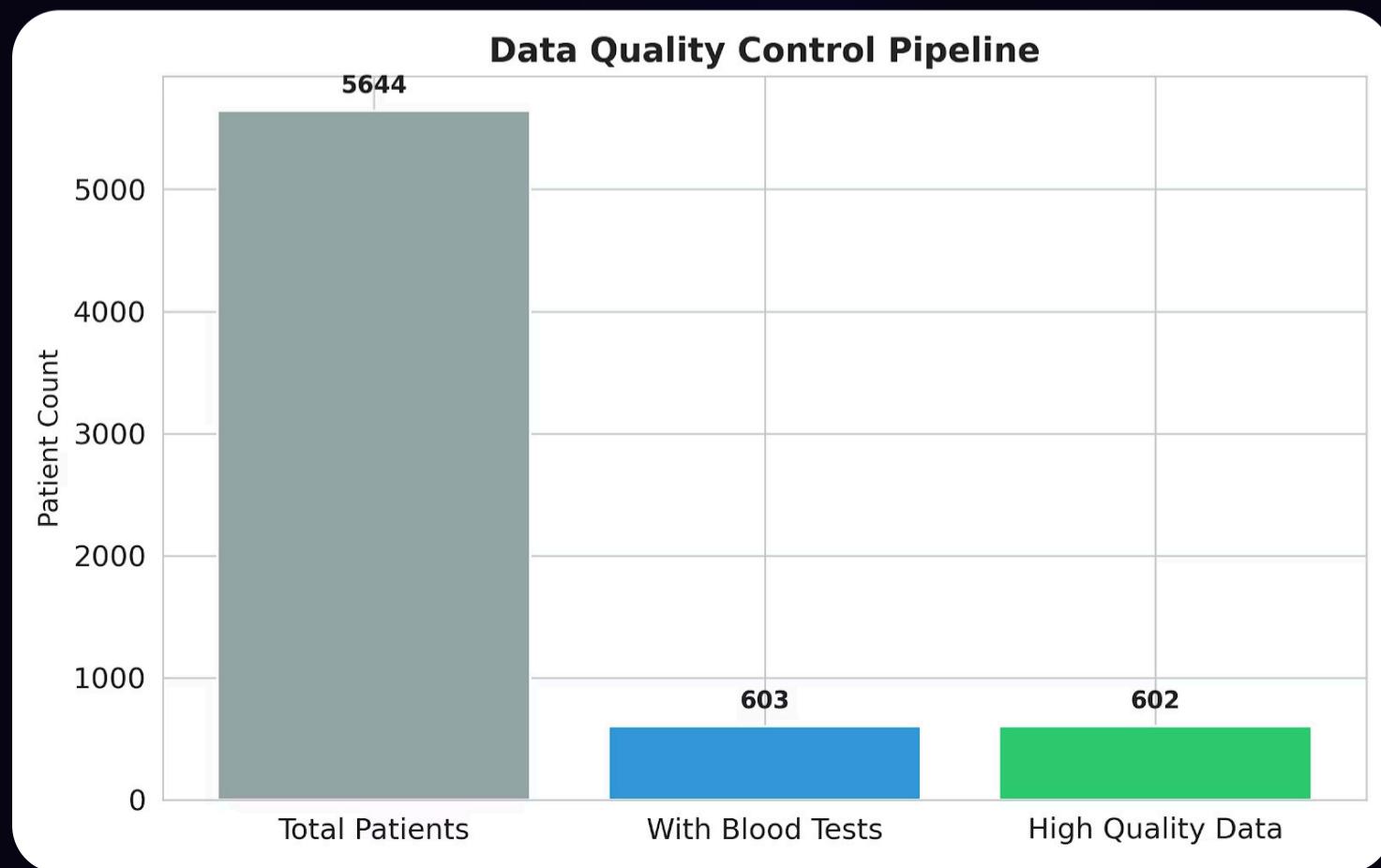
סוג הדטה: נתונים טבלאיים (Tabular Data).

פתרונות

- **סינון קליני קפדי:** יצרנו משפר סינון שהshore רק נבדקים עם פרופיל המטולוגי ובויכימי מלא. נותרנו עם 309 נבדקים איקוטיים.
- **טיפול נקודתי בחוסרים:** השלמה מינימלית באמצעות חציון רק על האוכלוסייה הרלוונטית, למניעת הטיאות.
- **נרמול:** ביצענו Scaling Score-Z כדי לאפשר השוואה הוגנת בין סוגי תאים שונים (למשל, מיליון כדריות אוזניות מול אלף כדריות לבנות).

האתגר

- סוג הדטה שלנו הוא נתונים טבלאיים ומסד הנתונים הנומלמי שלנו הכיל 5,644 נבדקים, אך רובם המוחלט עברו בדיקת מטוש (PCR) בלבד ללאלקיחת דם. ניתוך כלל המאגר היה מוביל לשימוש בנ נתונים מסונתזים ולא אמיתיים, מה שהיה פוגע באמינות הבiology של התוצאות.



מנתוניים גולמיים לתובנות קליניות

1

2

3

Unsupervised Learning

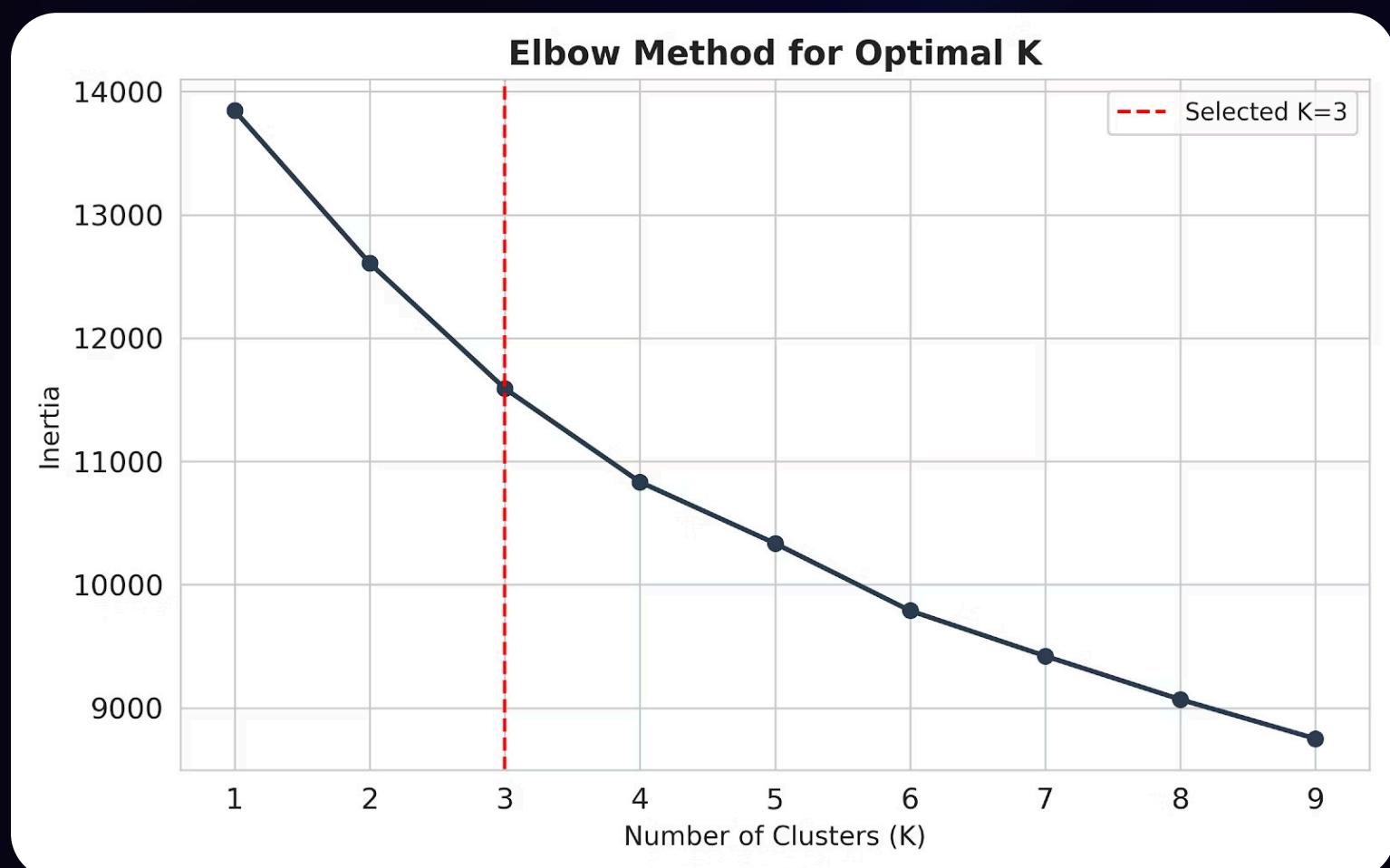
"שימוש אלגוריתם K-Means ליזיהו"
אשכולות(Clusters) נסתרים,
לאחר נירמול הנתונים (Z-Score)
למניעת עיוות מרחקים,
וללא שימוש בתוצאות מוקדמות.

Dimensionality Reduction

שימוש ב-PCA ליזיהוק המידע
(Feature Extraction) והמחשה
ווייזואלית של המרחב הפונטיפי
בזווית מינימום.

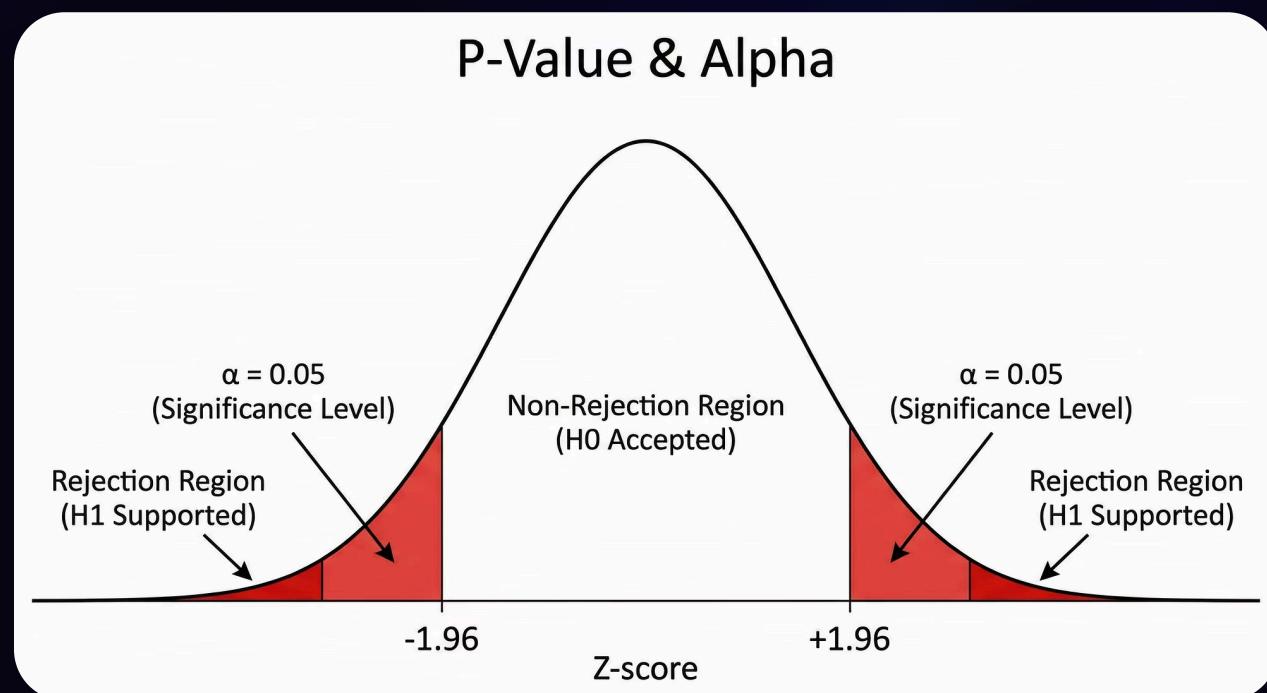
בחירה ב-K=3

בעוד ש- $K=2$ הפריד בין בריאות
לחולים, הבחירה ב- $K=3$ מאפשרת
למודל **לזהות** שונות פנימית
בתוך קבוצת החולים ולחשוף
תתי-דפוסים של המחלה.



השערת המחבר (Hypothesis)

1. **השערת האפס (H_0)**: לא קיים הבדל מובהק בממוצעי הבiomarkerים בין האשכבות(Clusters) שזויה.
2. **השערת החוקר (H_1)**: קיימים פנוטיפים קליניים מובהקים המאפיינים בפרופילים ביולוגיים שונים (למשל: דלקתי vs כשל איברים).
3. **אסטרטגיית התקופה**: דחינת השערת האפס באמצעות מבחנים סטטיסטיים מחמירים להוכחת שונות ביולוגית וקלינית בין הקבוצות.



אפיון הקבוצות

אשכול 2 - כשל רב מערכתי

המצאים: חריגת בולטת במזדים מטבוליים וכלייתיים המעידת על עומס פיזיולוגי קיצוני.

המשמעות: מצב קליני מורכב המצביע על פגיעה בתפקוד מערכות הסיכון - מנבא חזק להידררות וטיפול נמרץ

אשכול 1 - הפטוטיפ הדלקתי

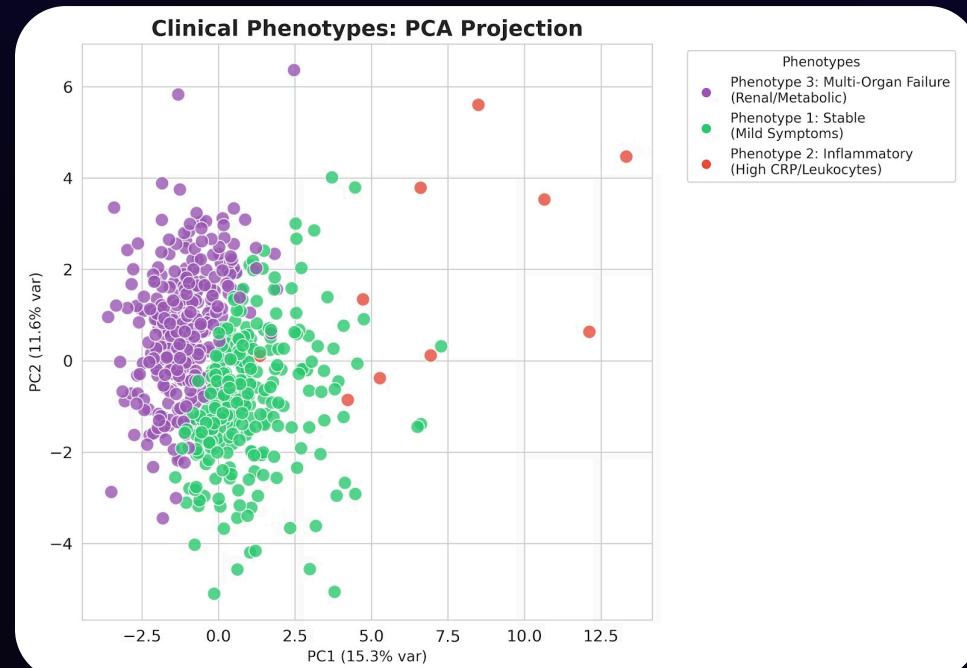
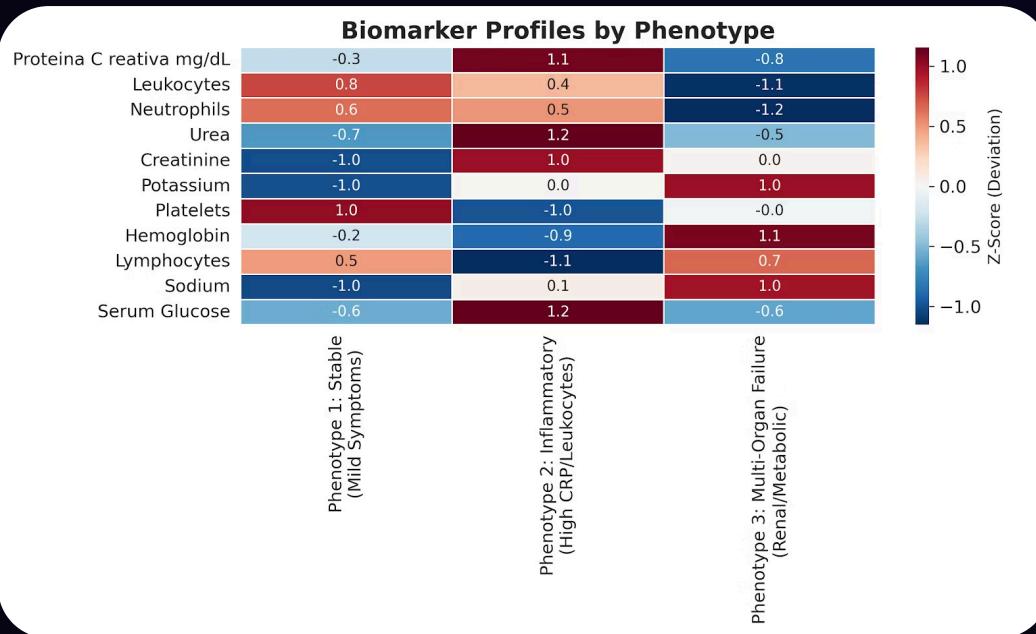
המצאים: עלייה עקבית במזדי דלקת (בעיקר CRP ולוקוציטים) המעידת על עומס אימונולוגי.

המשמעות: תגובה דלקתית חזקה ("סרת ציטוקינים"), אך ללא עדות לקיום איברים פנימיים בשלב זה.

אשכול 0 - הפרוfil היציב

המצאים: רוב המזדים קרובים לטווחים הרגילים, כולל מזדי דלקת נמוכים.

המשמעות: הקבוצה מייצגת את הפרוfil היציב ביותר בשלב הכניסה למיון. הוגן מכיל את המחלות ללא מצוקה מערכתיות.



השלבים הבאים

הוכחת היתכנות: בדיקות דם שגרתיות מאפרחות זיהוי ראשוני של קבוצה עם פרופיל חריג כבר בשלבי המין.

השלבים הבאים:

03

תיקוף חיצוני (Validation)

הערכת המודל על נתונים מבית חולים אחר, כדי לוודא שה��וסים שנ גילינו הם אוניברסליים ולא ייחודיים רק למדגם הנוכחי.

02

אופטימיזציה פיז'רים (Feature Optimization)

שימוש ב-Feature Importance כדי לאתר את 5-3 המזדיינים הクリティים ביותר, וליצור בדיקת בזק הסכונית ומתחזקת.

01

מעבר ללמידה מפוקחת (Supervised Learning)

פיתוח מודל חיזוי שיקבל בדיקות דם בזמן אמת ויסווג את המטופל לאחת מהקבוצות, ככלי תומך החלטה במין.

תודה רבה!

שאלות?

יוניברסיטת תל אביב | COVID-19 Clinical Phenotyping Research