נתונים והגדרות שאלת מחקר

נתון

דוח נתונים ממדדים שונים ברכב חשמלי.[[1]](#footnote-1)

דוח הכולל 7 מדדים:

טמפרטורה

טמפרטורה חיצונית

לחות

זרם

מתח

רעידות

סיבובי מנוע

דוח מתויג לפי תקלות 0.1.2.3, כאשר 0 מוגדר פעולה תקינה.

נתונים בדוח בעלי 2-6 סגמנטים המופרדים באמצעות נקודה. הנתונים לפי אמורים להיות מופרדים, סגמנט ראשון נתון עיקרי, סגמנט שני נתון משני.

בניתוח יש להשתמש אך ורק במשתנים:

זרם Current

מתח Voltage

טמפרטורה Temperature

*הוגדר להתעלם משאר המשתנים*

יש למצא

נקודת/ות קריסה של המערכת

לבחור מודל חיזוי אשר יחזה אם הנתונים מצביעים שהמערכת נעה לקריסה = לכיוון נקודת/ות קריסה.

הנחות יסוד

הצורך בהנחות יסוד, הוא היעדר מידע אשר מגיע עם הנתונים:

סוג רכב

האם תצפיות בוצעו בזמן עוקב

האם תצפיות בוצעו מאותו הרכב

מהי הגדרת היצרן לנקודות קריסה

מהי הגדרת היצרן לרמת קריטיות של תקלות 1/2/3

מה הם התנאים התקינים לפעילות המערכת (זרם, טמפרטורה,מתח) ומהי אנומליה.

מאילו חומרים פיזיים מורכבת המערכת (סוגי מתכות, סוגי פלסטיק וכד')

לפי כך הנחות היסוד לבדיקה ומחקר הן:

רכיבים פיזיים מהם מורכבת המערכת:

סגסוגת אלומיניום

סגסוגת ניקל

פלסטיק (חומר בעל נקודת היתוך הנמוכה ביותר מבין השלושה)

אנומליות

בהיעדר הגדרות אנומליות או לחלופין הגדרות הנורמה, נבחרו טווחי האנומליות הבאים:

Current Main > 250[[2]](#footnote-2)

Temperature > 250[[3]](#footnote-3)

Voltage > 850[[4]](#footnote-4)

נקודות קריסה

בהיעדר נתונים מנחים מהיצרן ובהינתן תיוג פעולה תקינה ולא פעולה תקינה, מניחים שכל אחת מהתקלות המתויגות כ 1,2,3 יכולה להיות תקלה שתגרום לקריסה ולכן אפשר להתייחס לנתונים כתקינים: בעלי תיוג 0 ו תקלה בעלי תיוג 1,2,3.[[5]](#footnote-5)

ניתוח נתונים

הנתונים מכילים 11K תצפיות, מתוכן 5K מתויגים כתקינים (Fault\_label=0) ו 6K מתויגים כתקלה (Falut\_Label >0 && Falut\_Label =< 3). ראה Table 1.

חלוקת הנתונים לשני קבצים תוך ביצוע חלוקה בכל קובץ ובכל עמודה לערך ראשי ומשני Equation 1

בחלוקה זוהו שתי תצפיות בעלות ערך תקול (לא נומרי), התצפיות בוטלו ונמחקו. Table 2

כתוצאה התקבלו שני בסיסי נתונים (קבצי csv)[[6]](#footnote-6):

נתונים תקינים (Fault\_label=0) ונתונים מתקלות (Falut\_Label >0 && Falut\_Label =< 3). לאור החלוקה הדיכוטומית בוטלה עמודה Fault\_Label בכל אחד מהקבצים.

נעשה סינון (ביטול תצפיות בעלות ערכים אנומליים Equation 1. דבר זה נועד לפני הכל "להציל" את מירב הנתונים ולשמור על היקף התצפיות קרוב לכל הניתן למקור ו/רוק חב ככל הניתן. הסיבה לביטול/מחיקת התצפיות המכילות ערכים אנומליים היא ב:

בוצע חישוב SVD, תיוג נתונים אנומליים כ NA והשלמת הנתונים המתוייגים כ NA בערכים מנורמלים באמצעות מודל SVD בפועל החלפת הערכים החסרים NA בערכים החזויים. הדבר אינו עמד במבחן ה MSE, הריבועים של ההפרשים בין הערכים החזויים לערכי אמת (rowSet שנתון לנו).

היעדר דיוק בהמרת תצפיות אנומליות בערכים מנורמליים[[7]](#footnote-7)

כתוצאה התקבלו נתונים מסוננים :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Fault 0 | | | | | |
|  | mean | median | std | min | max | count |
| Voltage\_main | 43.7 | 8 | 161.9 | 7 | 850 | 4150 |
| Voltage\_secondary | 548.6 | 578 | 285.7 | 0 | 999 | 4150 |
| Current\_main | 12.9 | 7 | 16.8 | 1 | 248 | 4150 |
| Current\_secondary | 501.2 | 503 | 288.4 | 0 | 999 | 4150 |
| Temperature\_main | 16.1 | 8 | 20.1 | 1 | 245 | 4150 |
| Temperature\_secondary | 486.8 | 483 | 283.8 | 0 | 999 | 4150 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Fault 1-3 | | | | | |
|  | mean | median | std | min | max | count |
| Voltage\_main | 53.5 | 7 | 164 | 0 | 850 | 5181 |
| Voltage\_secondary | 490.4 | 479 | 289.7 | 0 | 999 | 5181 |
| Current\_main | 15.3 | 7 | 19.7 | 0 | 250 | 5181 |
| Current\_secondary | 495.3 | 500 | 283.4 | 0 | 999 | 5181 |
| Temperature\_main | 13.9 | 7 | 19 | 0 | 248 | 5181 |
| Temperature\_secondary | 495.2 | 494 | 283.7 | 0 | 999 | 5181 |

לנתונים אין קורלציה בין המשתנים (Figure 1, Figure 2) או שקורלציה מאוד חלשה (לרוב סביב המאית).

שלב נוסף בוצע ניתוח PCA (למה לא SVD? מהטעם הפשוט שרוצים לזהות קבוצות של משתנים שמשפיעות וקלות של הצגה, PCA ו SVD הינם מודלים/אופני עיבוד/שיטות קרובים אחד לשני, אך שונים. מכיוון שרצוננו לזהות את הקשרים ואת קבוצת המשתנים המשפיעה ומה מתשמעותה אחוזית על השונות נבחר PCA.

בניתוח PCA אפשר לזהות שאין משתנה אחד או קבוצה אחת אשר מסבירים את מירב השונות (Figure 3, Figure 4) ויש צורך בלמעלה מארבע קבוצות על מנת להסביר מעט מעל 80% מהשונות

(Figure 5, Figure 6).

לכן על מנת לזהות את הנקודות הקריטיות יש לבחור בניתוח השכיחויות. לנתח את המירב (75%) הנתונים לפי שכיחות אשר מתויגים כתקלה ואלה שמתויגים כתקינים. מכיוון שזוהתה השפעה משמעותית של משתנה temperature\_secondary 0.88 ב PC4 (Fugure 4 לשנות) והמשתנה לא בא לידי ביטוי בתנאי עבור זיהוי נקודות קריטיות, קיימת הבנה שהתעלמות מהמשנה הזה מחוייבת הצדקה רציונלית ומבוססת.

ערכים משניים, או, משתניים משניים, בהחלט מתנהגים אחרת מהמשתנים הראשיים. (Figure 13) בניגוד למשתנים ראשיים, הם, המשתנים המשניים קרובים יותר להתפלגות נורמלית וזה מעורר תהייה מה ערכם בפועל מייצג. **בהיעדר הסבר מהו הערכך המשני** (חלק ממדידת הערך הראשי, אם כן, ההתנהגות האחרת שוללת את זה, האם זה בכלל לא הערך שקשור לשמתנה, אולי זו נקודת זמן, או משהו שהוא לא Voltage, במקרה של מתח. בהיעדר הסבר היצרן/לקוח/עורך התצפיות’ משתנים משניים הם נעדרי משמעות.

ולכן אפשר לראות בבירור ש:

¾ מנתוני Voltage Main מרוכזים בין 7V – 9V במצב תקין וכ 45% מהנתונים המתוייגים כתקלה , במתח, מרוכזים בין 0V-6V. (Figure 8, Figure 7)

כ 53% מהנתונים המתוייגים כתקיניים, בזרם, מרוכזים בין 5 -9. בנתונים המתוייגים כתקולים כ 30% מרוכזים בין 0 I – 5 I וכ 15% מעל 9 I. (Figure 10, Figure 9)

עיקר מדדי הטמפרטורה בנתונים תקינים עד 5C ובנתונים המתוייגים כשגויים מעל 5C. (Figure 12, Figure 11)

לכן נקודות הקריסה שנבחרו הן:

crash =

(temp.main > 5.8 AND (current.main > 5 OR current.main < 4) AND voltage.main < 7)

נקודות אלה מאפשרות לזהות 26.29% ממקרי תקלה עם 0% זיהוי כושל של תקלה מתוך נתונים שתויגו בתקינים. Equation 3

מודל לחיזוי התנהגות שמתקרבת או הולכת אל התקלה שנבחר הוא prophet בעיקר בגלל יכולת לעבוד עם סדרות זמן ונתונים לא לניאריים. (חלופה שנשקלה Arima model).[[8]](#footnote-8)

לשם כך נבנו קבצי/בסיסי נתונים עם ערך עולה בכל תצפית בנשיה אחת.

נספחים

Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Count | | |
| Fault Label | Voltage (V) | Current (A) | Temperature (C) |
| 0 | 5000 | 5000 | 5000 |
| 1 | 2000 | 2000 | 2000 |
| 2 | 2000 | 2000 | 2000 |
| 3 | 2000 | 2000 | 2000 |

Equation

xxx.yyy.zzz.hhh.www.kkk

xxx

סגמנט ראשון – נתון עיקרי

yyy

סגמנט שני – נתון משני

Table

|  |  |  |  |
| --- | --- | --- | --- |
| Fault Label | Voltage (V) | Current (A) | Temperature (Â°C) |
| 2 | 5.542.896.479.769.260 | 8,03E+10 | 24.911.848.629.758.800 |
| 2 | 5.045.887.171.680.880 | 6,22E+10 | 28.416.670.806.151.000 |

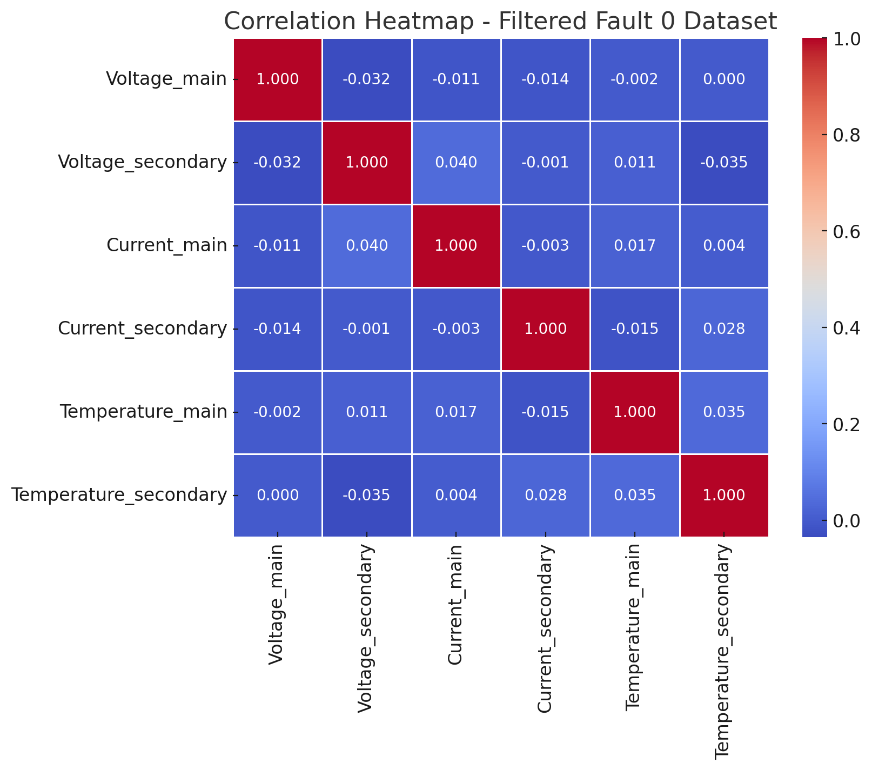
Equation

Current Main > 250

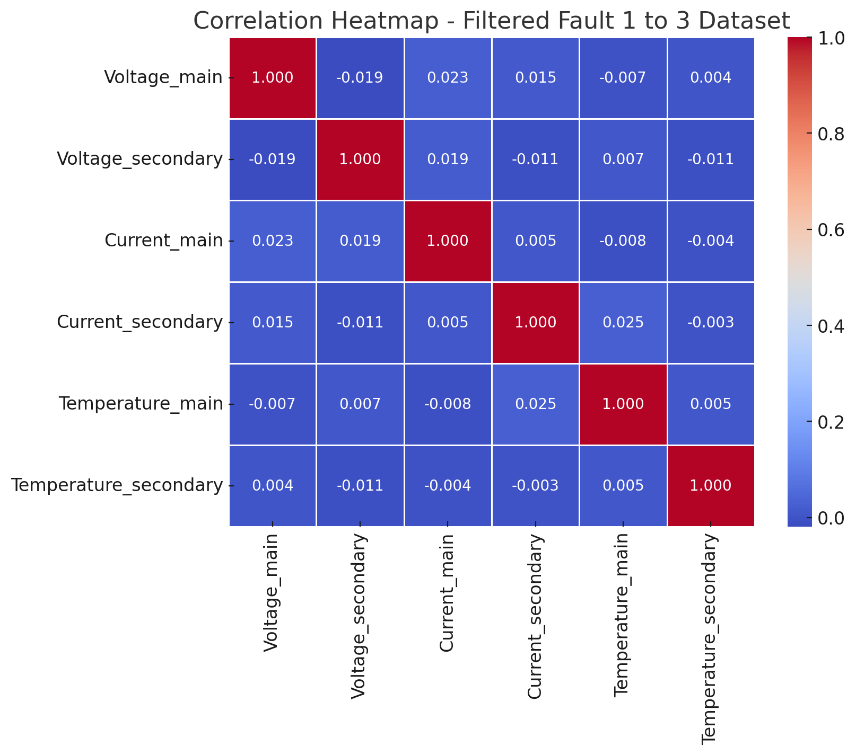
Temperature > 250

Voltage > 850

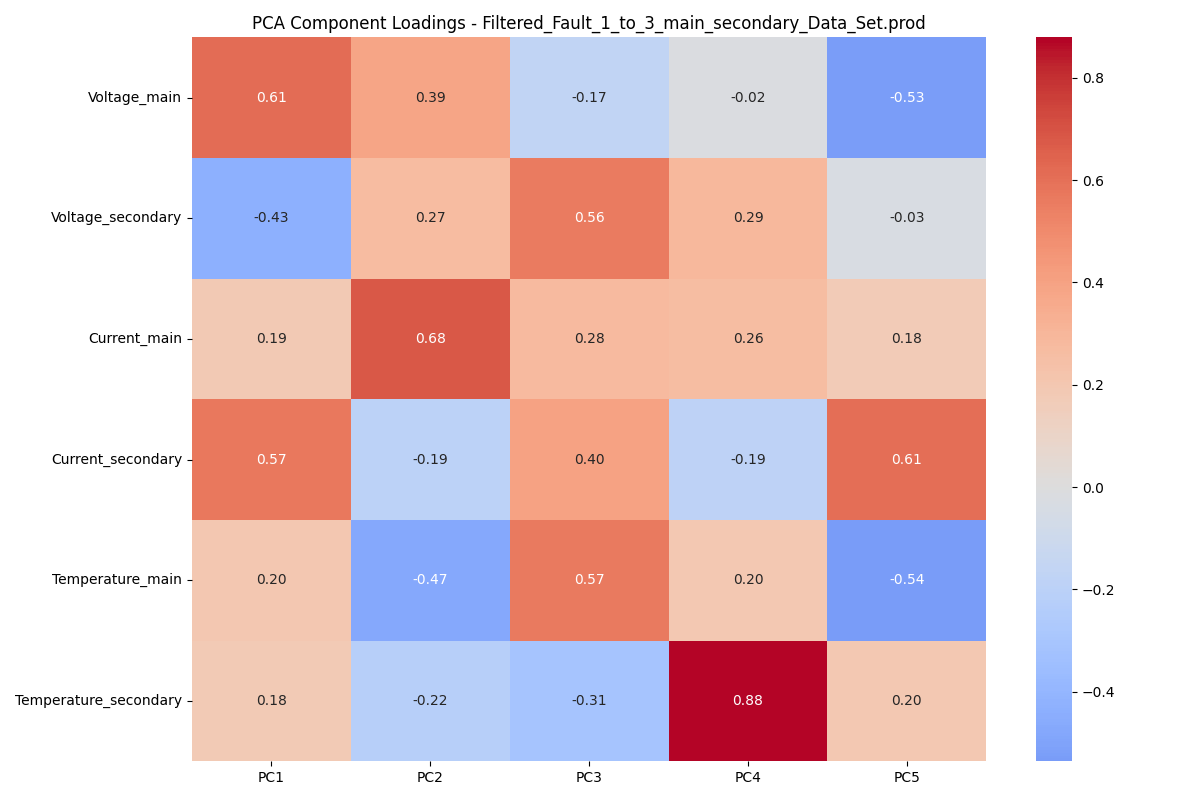
Figure



Figure



Figure

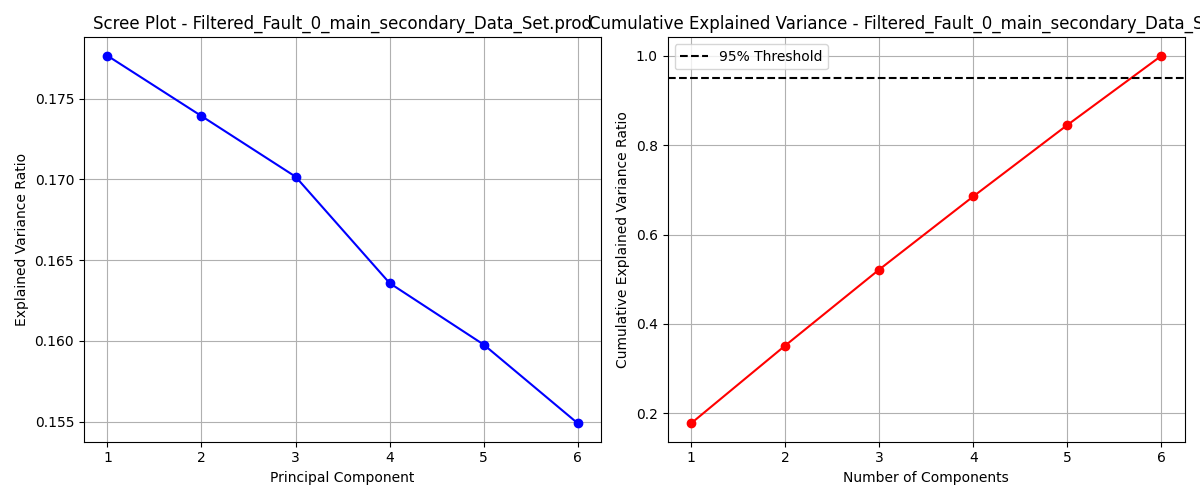


Figure

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure

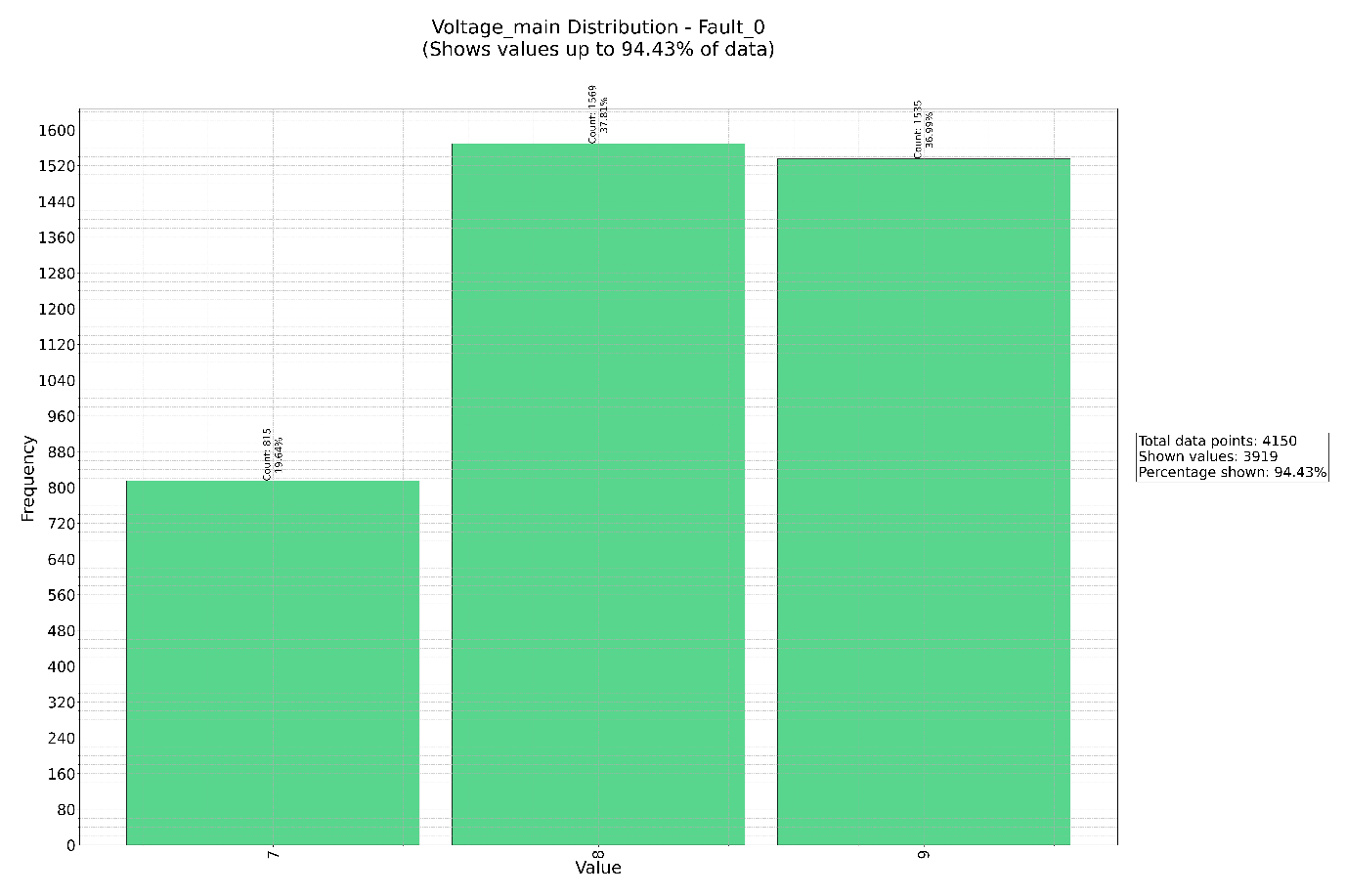


Figure

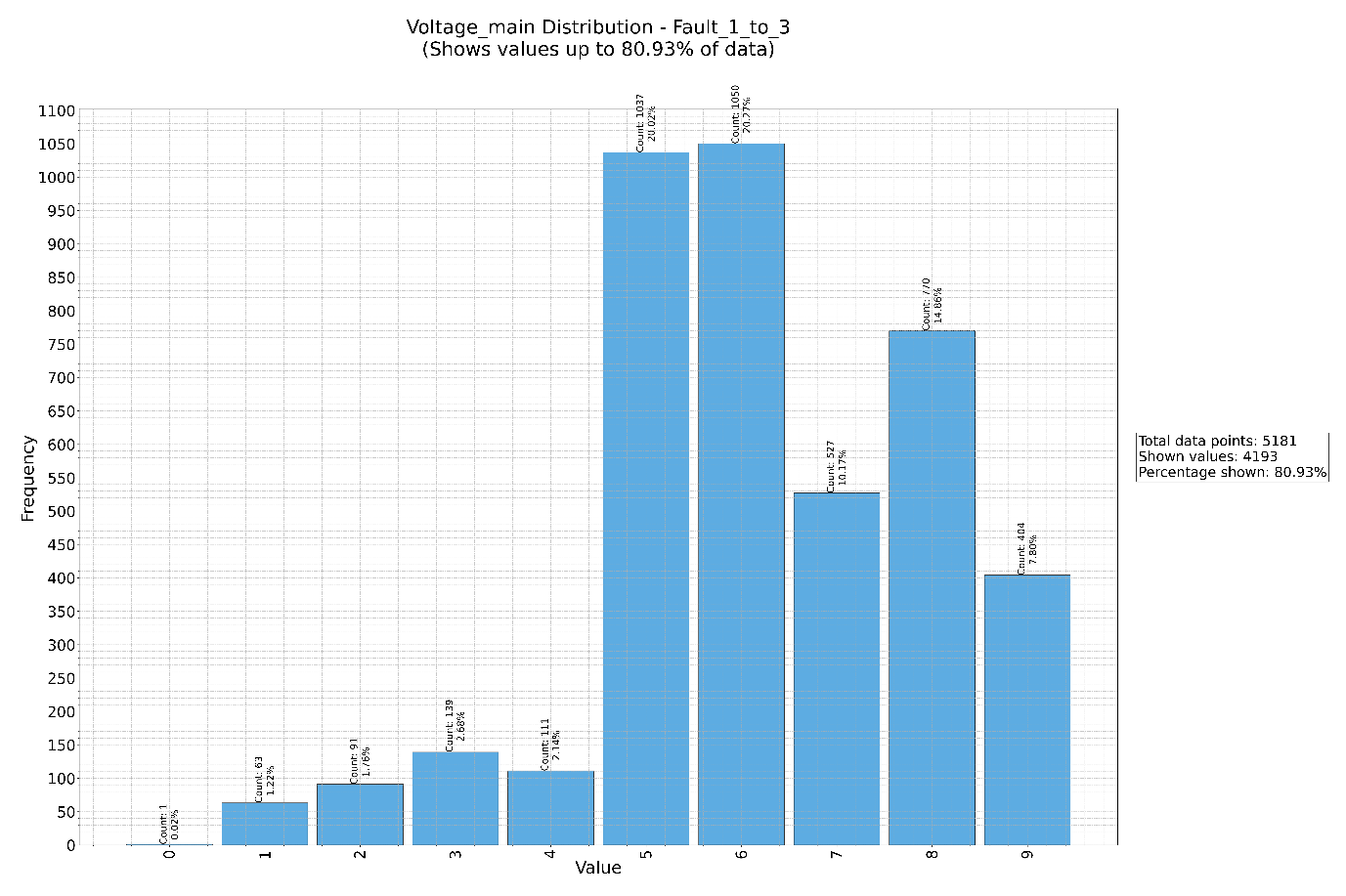
A graph of a line and a line

AI-generated content may be incorrect.

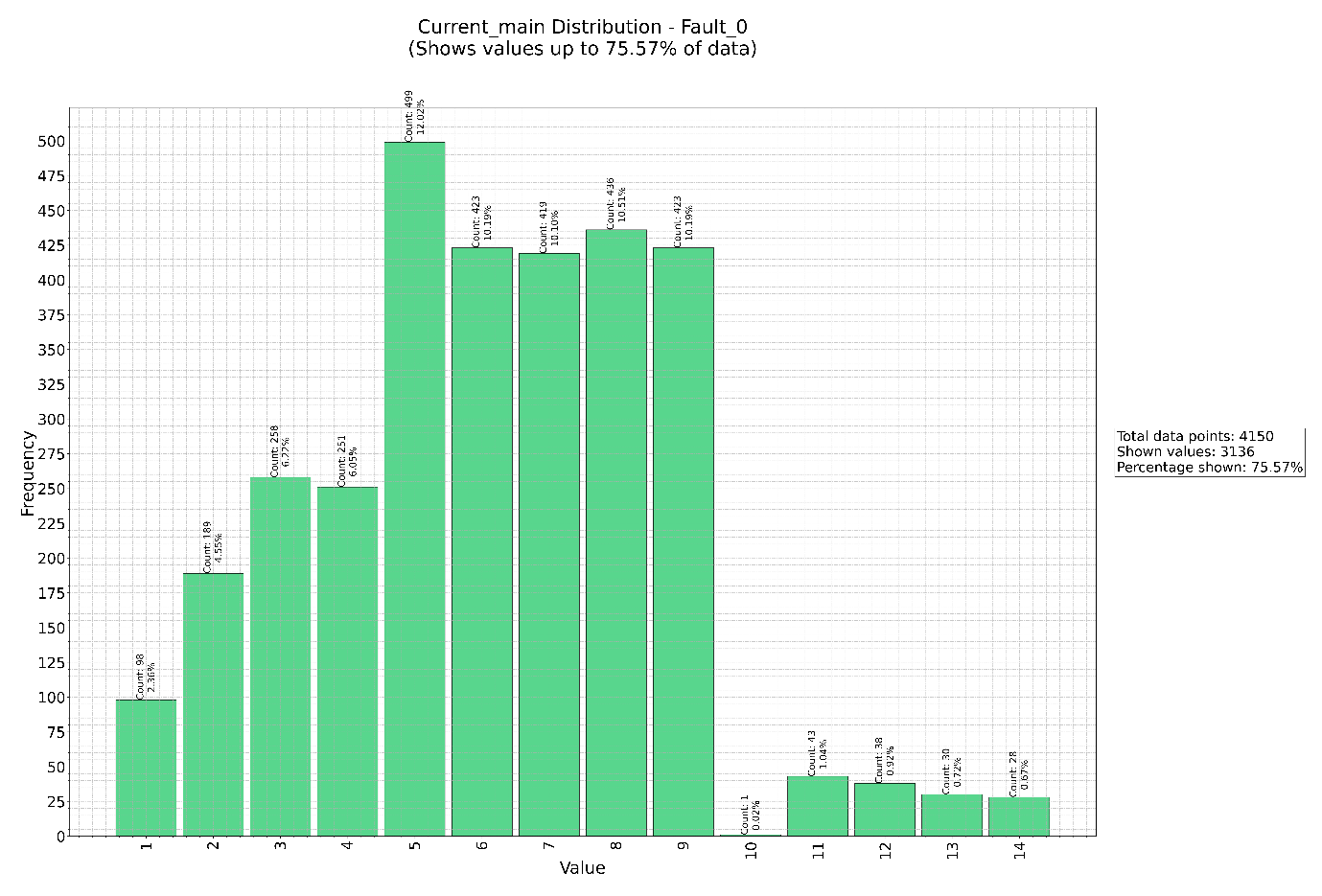
Figure



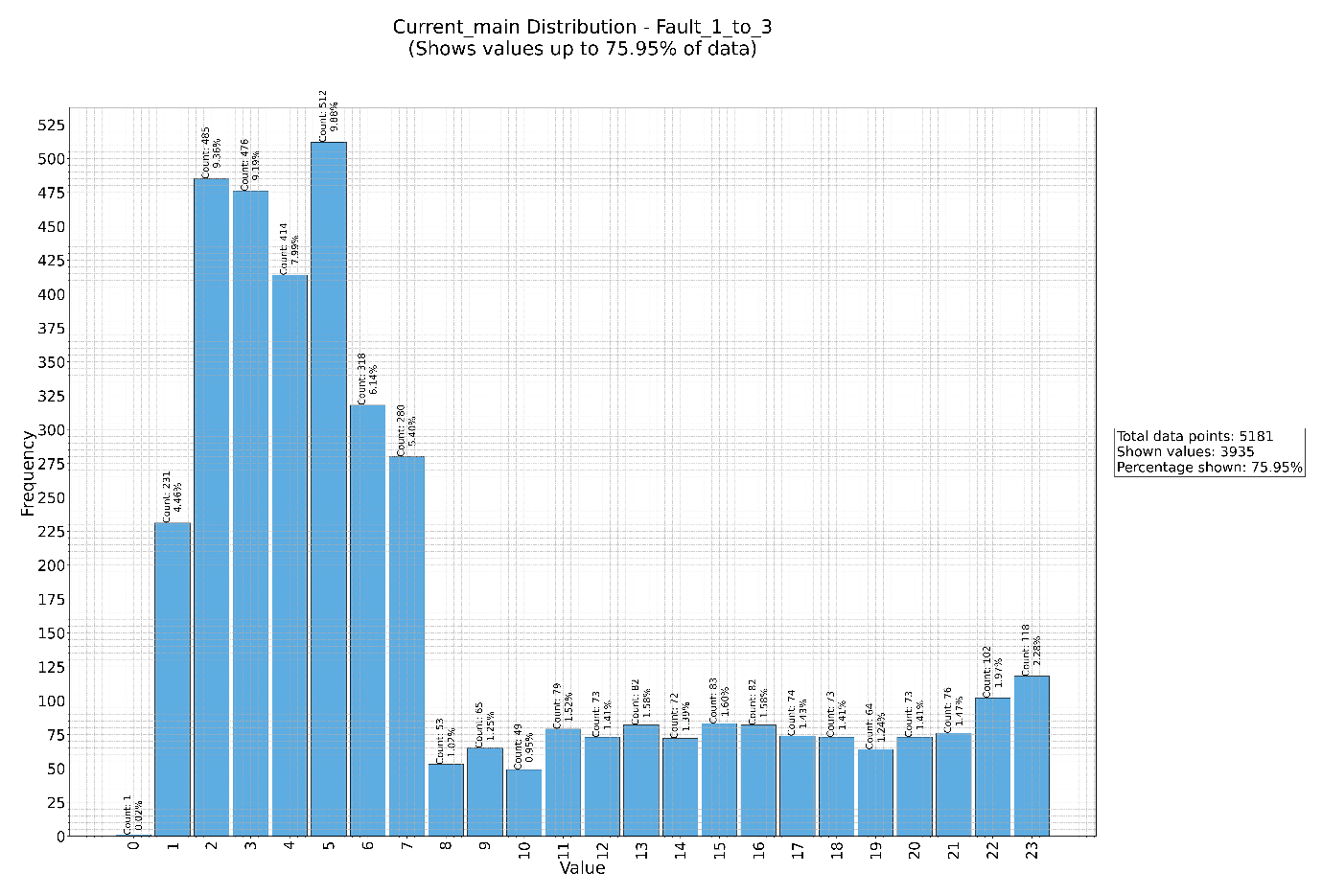
Figure



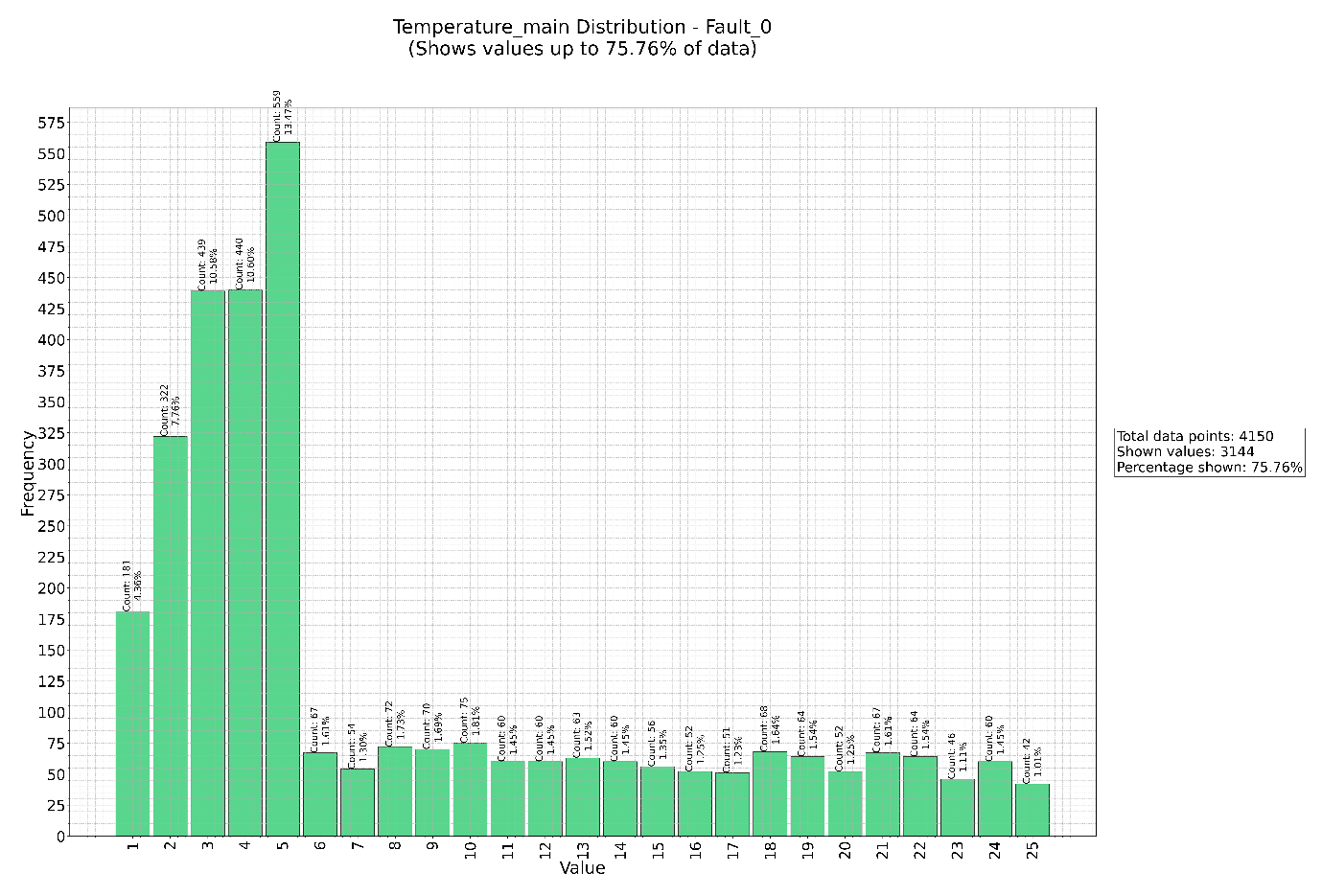
Figure



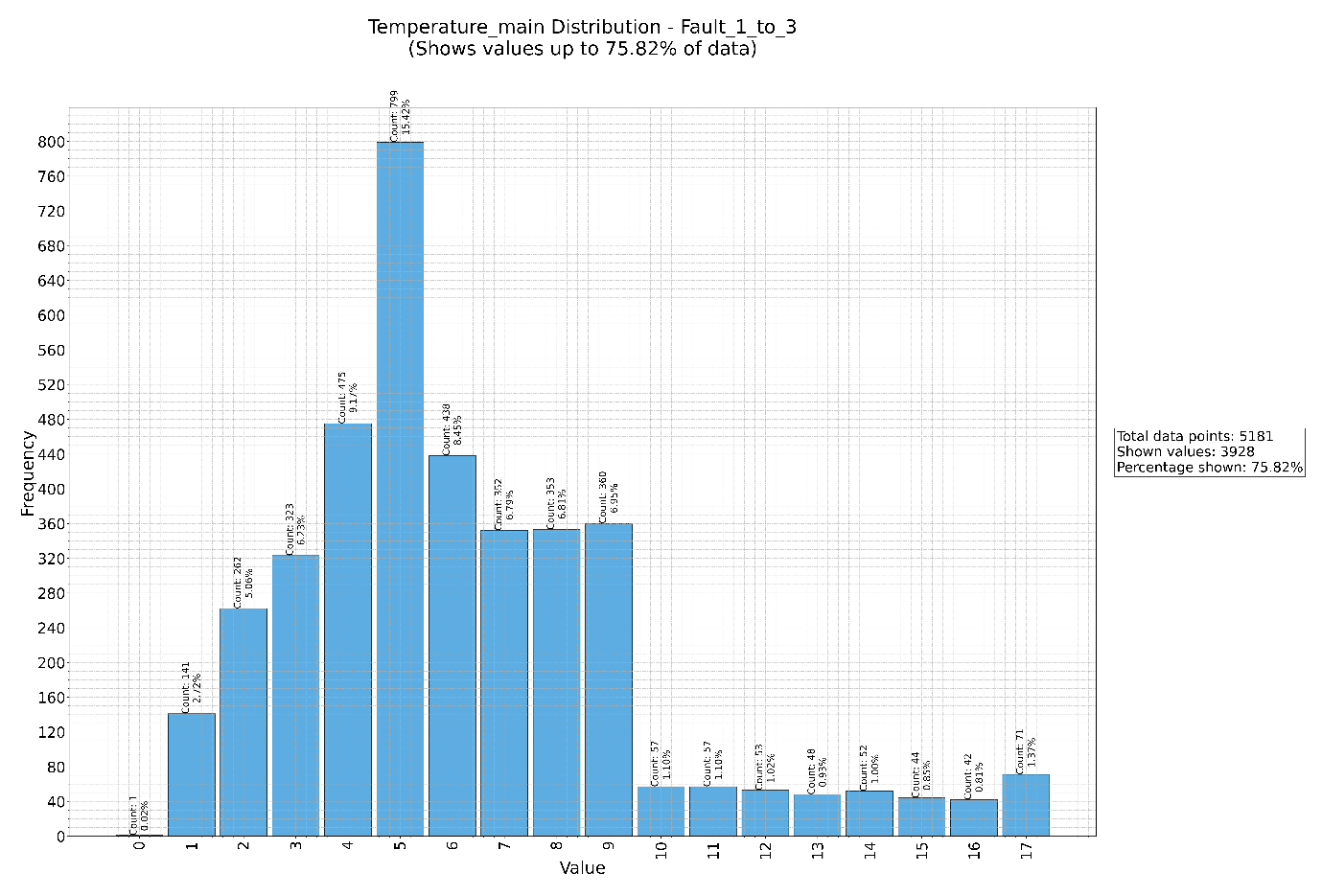
Figure



Figure



Figure



Equation

Fault Detection Rate (26.29%):

Total fault cases in dataset: 5,181

Number of faults detected by your method: 1,362

Calculation: (1362/5181) × 100 = 26.29%

False Positive Rate (0%):

Total normal cases: 4,150

Number of false alarms: 0

Calculation: (0/4150) × 100 = 0%

Calculation: 1362/(1362 + 0) × 100 = 100%

Precision (100%):

True Positives: 1,362

False Positives: 0

Figure

A screenshot of a graph

AI-generated content may be incorrect.

1. <https://github.com/polonskyilya/Data/blob/main/Prod/CSV_DATA_SETS/Original_Data_Set.prod.csv> [↑](#footnote-ref-1)
2. power (Watts) is calculated by multiplying voltage (Volts) by current (Amps). With 250 I and 850 V are estimated to have 212.5 KW which could be equal to 284Hp (quite extreme for EV).

   [https://brainly.in/question/40317349#:~:text=How%20much%20electricity%20s%20consumed,will%20use%203728.5%20kilowatt%2Dhours.](https://brainly.in/question/40317349" \l ":~:text=How%20much%20electricity%20s%20consumed,will%20use%203728.5%20kilowatt%2Dhours) [↑](#footnote-ref-2)
3. על בסיס נקודות התכה בפלסטיק

   חומר בעל נקודת ההתכה הנמוכה ביותר ברכב  
   <https://www.plastikcity.co.uk/useful-stuff/material-melt-mould-temperatures> [↑](#footnote-ref-3)
4. על בסיס מבנה סוללות לרכבים חשמליים

   <https://www.power-sonic.com/blog/400v-vs-800v-charging/#:~:text=Today%2C%20most%20EVs%20are%20built,higher%2Dvoltage%20battery%20can%20offer> [↑](#footnote-ref-4)
5. Fault label value in data set <https://github.com/polonskyilya/Data/blob/main/Prod/CSV_DATA_SETS/Original_Data_Set.prod.csv> [↑](#footnote-ref-5)
6. <https://github.com/polonskyilya/Data/blob/main/Prod/CSV_DATA_SETS/Main_Secondary_DataSet/Fault_1_to_3_main_secondary_Data_Set.prod.csv>

   <https://github.com/polonskyilya/Data/blob/main/Prod/CSV_DATA_SETS/Main_Secondary_DataSet/Fault_0_main_secondary_Data_Set.prod.csv> [↑](#footnote-ref-6)
7. ניסיון למלא את הנתונים האנומליים בנתונים מנורמלים באמצעות SVD הניב תוצאות בעיתיות בעקבות בדיקת MSE לדוגמא

   Current measurements: Main: 34561.928010 Secondary: 0.000000

   Temperature measurements: Main: 13199.103741 Secondary: 0.000000 [↑](#footnote-ref-7)
8. <https://medium.com/@tarangds/traditional-prediction-models-prophet-arima-83bc8b980ec4> [↑](#footnote-ref-8)