



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

Duomenų tyryba Pirminis duomenų aibės apdorojimas.

Savarankiško darbo ataskaita

Atliko: Justinas Rimavičius, Edvardas
Ražanskas

VU el. p.: edvardas.razanskas@mif.stud.vu.lt,
justinas.rimavicius@mif.stud.vu.lt

Vertino: dr. Jolita Bernatavičienė

Vilnius
2024

1. Įvadas	3
1.1. Darbo tikslas	3
1.2. Darbo uždaviniai.....	3
2. Duomenų analizė.....	5
2.1. Tiriamos duomenų aibės ir jos požymių aprašymas.....	5
2.2. Duomenų aprašomoji statistika (min, max, 1, 3 kvartilės, vidurkis, mediana, standartinis nuokrypis).	7
2.3. Praleistų reikšmių pildymas, šalinimas.....	8
2.4. Taškai atsiskyrėliai.....	9
2.5. Duomenų aibės normavimas.	10
2.6. Duomenų aibės vizualizacija, analizė, išvados.	11
2.7. Požymių tarpusavio priklausomybės analizė.....	19
3. Išvados.....	27
4. Priedai.....	Error! Bookmark not defined.
4.1. Kodas.....	Error! Bookmark not defined.

1. ĮVADAS

1.1. Darbo tikslas

Šio darbo tikslas yra atlikti pateiktos EKG pūpsnių duomenų aibės pirminį apdorojimą, išsamią analizę taikant statistinius metodus ir duomenų vizualizaciją, siekiant išnagrinėti požymių pasiskirstymą, tarpusavio priklausomybę bei paruošti duomenis tolimesniems tyrimams ar modeliavimo uždaviniams.

1.2. Darbo uždaviniai

1. Trumpai aprašyti tiriamą duomenų aibę, jos požymius, pagrindines savybes.
2. Pateikti atskirų požymių aprašomąsias statistikas.
3. Užpildyti ar ištrinti praleistas reikšmes.
4. Sunormuoti duomenų aibę.
5. Pateikti vizualią duomenų aibės analizę, aprašyti rezultatus.
6. Iširti požymių tarpusavio priklausomybę.

1.3 ĮRANKIAI: pytyhon ir python bibliotekos; 2.7 pav. ir aiskus komentarai prie kodo.

2. DUOMENŲ ANALIZĖ

2.1. Tiriamos duomenų aibės ir jos požymių aprašymas.

Pateiktoje duomenų aibėje yra 11940 eilutės, 32 požymių stulpeliai. Dauguma reikšmių yra „float“ tipo, keli įrašai – „object“(‘a’, ‘#NAME?’ skaitomi kaip objektai).

Pateikiami EKG duomenys:

RR intervalai kairėje taško pusėje: RR_1_0, RR_1_1, RR_1_2, RR_1_3, RR_1_4.

RR intervalai dešinėje atskaitos taško pusėje: RR_r_0, RR_r_1, RR_r_2, RR_r_3, RR_r_4

Kairiųjų RR intervalų santykiai: RR_1_0/RR_1_1, RR_1_1/RR_1_2, RR_1_2/RR_1_3, RR_1_3/RR_1_4

Dešiniųjų RR intervalų santykiai: RR_r_0/RR_r_1, RR_r_1/RR_r_2, RR_r_2/RR_r_3, RR_r_3/RR_r_4

Sekos dydis: seq_size

Signalų vidurkis: signal_mean

Signalų standartinis nuokrypis: signal_std

P, Q, R, S, T EKG bangų amplitudės: P_val, Q_val, R_val, S_val, T_val

P, Q, R, S, T EKG bangų pozicijos: P_pos, Q_pos, R_pos, S_pos, T_pos

Kairiosios pusės ilgis: wl_side

Dešinėsios pusės ilgis: wr_side

Klasės žymė: label.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11940 entries, 0 to 11939
Data columns (total 32 columns):
#   Column                Non-Null Count  Dtype
---  -
0   RR_l_0                 11940 non-null  float64
1   RR_l_0/RR_l_1         11939 non-null  float64
2   RR_l_1                 11939 non-null  float64
3   RR_l_1/RR_l_2         11938 non-null  float64
4   RR_l_2                 11938 non-null  float64
5   RR_l_2/RR_l_3         11926 non-null  float64
6   RR_l_3                 11936 non-null  float64
7   RR_l_3/RR_l_4         11933 non-null  float64
8   RR_r_0                 11932 non-null  object
9   RR_r_0/RR_r_1         11936 non-null  object
10  RR_r_1                 11935 non-null  float64
11  RR_r_1/RR_r_2         11932 non-null  float64
12  RR_r_2                 11937 non-null  float64
13  RR_r_2/RR_r_3         11935 non-null  float64
14  RR_r_3                 11938 non-null  float64
15  RR_r_3/RR_r_4         11938 non-null  float64
16  seq_size               11938 non-null  float64
17  signal_mean           11935 non-null  float64
18  signal_std            11937 non-null  float64
19  wl_side               11939 non-null  float64
20  wr_side               11936 non-null  float64
21  P_val                 11938 non-null  float64
22  Q_val                 11935 non-null  float64
23  R_val                 11935 non-null  float64
24  S_val                 11937 non-null  float64
25  T_val                 11939 non-null  float64
26  P_pos                 11940 non-null  float64
27  Q_pos                 11938 non-null  float64
28  R_pos                 11940 non-null  float64
29  S_pos                 11938 non-null  float64
30  T_pos                 11940 non-null  float64
31  label                 11923 non-null  float64
dtypes: float64(30), object(2)
memory usage: 2.9+ MB
```

Pav 2.1. Pradiniai duomenys prieš tvarkymą

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11940 entries, 0 to 11939
Data columns (total 32 columns):
#   Column                Non-Null Count  Dtype
---  -
0   RR_l_0                 11940 non-null  float64
1   RR_l_0/RR_l_1         11939 non-null  float64
2   RR_l_1                 11939 non-null  float64
3   RR_l_1/RR_l_2         11938 non-null  float64
4   RR_l_2                 11938 non-null  float64
5   RR_l_2/RR_l_3         11926 non-null  float64
6   RR_l_3                 11936 non-null  float64
7   RR_l_3/RR_l_4         11933 non-null  float64
8   RR_r_0                 11931 non-null  float64
9   RR_r_0/RR_r_1         11935 non-null  float64
10  RR_r_1                 11935 non-null  float64
11  RR_r_1/RR_r_2         11932 non-null  float64
12  RR_r_2                 11937 non-null  float64
13  RR_r_2/RR_r_3         11935 non-null  float64
14  RR_r_3                 11938 non-null  float64
15  RR_r_3/RR_r_4         11938 non-null  float64
16  seq_size               11938 non-null  float64
17  signal_mean           11935 non-null  float64
18  signal_std            11937 non-null  float64
19  wl_side               11939 non-null  float64
20  wr_side               11936 non-null  float64
21  P_val                 11938 non-null  float64
22  Q_val                 11935 non-null  float64
23  R_val                 11935 non-null  float64
24  S_val                 11937 non-null  float64
25  T_val                 11939 non-null  float64
26  P_pos                 11940 non-null  float64
27  Q_pos                 11938 non-null  float64
28  R_pos                 11940 non-null  float64
29  S_pos                 11938 non-null  float64
30  T_pos                 11940 non-null  float64
31  label                 11923 non-null  float64
dtypes: float64(32)
memory usage: 2.9 MB
```

Pav 2.2. Pradiniai duomenys po tvarkymo

Pradiniai duomenys(pav. 2.1) turėjo du stulpelius su „object“ tipo duomenimis, kas turėtų būti „float“ arba „int“. Sekančioje lentelėje(pav. 2.2) „object“ tipas yra pakeistas.

2.2. Duomenų aprašomoji statistika (min, max, 1, 3 kvartilės, vidurkis, mediana, standartinis nuokrypis).

```
target_cols = ['P_val', 'Q_val', 'R_val', 'S_val',
               'P_pos', 'Q_pos', 'R_pos', 'S_pos', 'signal_mean', 'label']
df[target_cols].describe()
```

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	signal_mean	label
count	11938.000000	11935.000000	11935.000000	11937.000000	11940.000000	11938.000000	11940.000000	11938.000000	11935.000000	11923.000000
mean	0.062501	-0.343119	1.322760	-0.482264	0.374371	0.440672	0.492280	0.545503	-0.199671	0.676340
std	0.452253	0.447397	0.592589	0.461104	0.100882	0.100863	0.097599	0.114927	0.142889	0.862264
min	-1.690184	-2.729307	-0.456913	-2.603065	0.055000	0.010000	0.060000	0.095000	-1.829613	0.000000
25%	-0.216751	-0.503635	0.850584	-0.595864	0.325000	0.435000	0.490000	0.515000	-0.271776	0.000000
50%	-0.117789	-0.329186	1.247799	-0.474172	0.355000	0.465000	0.500000	0.535000	-0.206050	0.000000
75%	0.199369	-0.188902	1.770605	-0.342989	0.405000	0.480000	0.515000	0.580000	-0.129889	2.000000
max	2.496907	2.658497	3.137667	2.699080	0.905000	0.870000	0.965000	0.995000	1.351685	2.000000

Pav 2.3. Pasirinktų požymių aprašomoji statistika

Tolimesniam apdorojimui pasirinkome P_val, Q_val, R_val, S_val, P_pos, Q_pos, R_pos, S_pos, signal_mean ir label požymius (pav. 2.3). Matoma visa reikiama informacija – 0,25, 0,75 kvartilų reikšmės, minimumo, maksimumo reikšmės, vidurkis, mediana, standartinis nuokrypis.

2.3. Praleistų reikšmių pildymas, šalinimas.

Viso identifikuotos 122 eilutės su praleistomis reikšmėmis iš 11940, t.y. maždaug 1.02178%.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11940 entries, 0 to 11939
Data columns (total 34 columns):
#   Column                Non-Null Count  Dtype
---  -
0   RR_l_0                 11940 non-null  Int64
1   RR_l_0/RR_l_1         11940 non-null  Float64
2   RR_l_1                 11940 non-null  Int64
3   RR_l_1/RR_l_2         11940 non-null  Float64
4   RR_l_2                 11940 non-null  Int64
5   RR_l_2/RR_l_3         11940 non-null  Float64
6   RR_l_3                 11940 non-null  Int64
7   RR_l_3/RR_l_4         11933 non-null  Float64
8   RR_l_4                 11933 non-null  Int64
9   RR_r_0                 11940 non-null  Int64
10  RR_r_0/RR_r_1         11940 non-null  Float64
11  RR_r_1                 11940 non-null  Int64
12  RR_r_1/RR_r_2         11940 non-null  Float64
13  RR_r_2                 11940 non-null  Int64
14  RR_r_2/RR_r_3         11940 non-null  Float64
15  RR_r_3                 11940 non-null  Int64
16  RR_r_3/RR_r_4         11938 non-null  Float64
17  RR_r_4                 11938 non-null  Int64
18  seq_size               11940 non-null  Int64
19  signal_mean           11935 non-null  Float64
20  signal_std            11937 non-null  Float64
21  wl_side               11940 non-null  Int64
22  wr_side               11940 non-null  Int64
23  P_val                 11938 non-null  Float64
24  Q_val                 11935 non-null  Float64
25  R_val                 11935 non-null  Float64
26  S_val                 11937 non-null  Float64
27  T_val                 11939 non-null  Float64
28  P_pos                 11940 non-null  Float64
29  Q_pos                 11938 non-null  Float64
30  R_pos                 11940 non-null  Float64
31  S_pos                 11938 non-null  Float64
32  T_pos                 11940 non-null  Float64
33  label                 11940 non-null  Int64
dtypes: Float64(20), Int64(14)
memory usage: 3.5 MB
```

Pav 2.4. Požymių aprašymas pakeitus jų tipus

Lentelėje (pav. 2.4) matome pagrindinę informaciją apie duomenis po nulinių reikšmių užpildymo ir tipų keitimo (iš „float“ į „int“ kur reikalinga).

RR_l_1, RR_r_1, RR_l_2, ... ir jų santykių pildymui naudojome tikslų pildymo būdą - išvedimą iš susijusių požymių.

Seq_size, wl_side ir wr_side pildymui taip pat naudojome išvedimą iš kitų reikšmių.

‘label’ pildymui naudojome forward fill metodą, kadangi duomenyse yra tik 3 unikalios reikšmės, ir jos eina viena paskui kitą.

Visas likusias eilutes, kuriose yra tuščių reikšmių, panaikinome (pav. 2.5). Tokių buvo tik 37, t.y. buvo panaikintas tik maždaug 0,32% eilučių.

```
<class 'pandas.core.frame.DataFrame'>
Index: 11903 entries, 0 to 11939
Data columns (total 34 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   RR_l_0                11903 non-null  Int64
1   RR_l_0/RR_l_1        11903 non-null  Float64
2   RR_l_1                11903 non-null  Int64
3   RR_l_1/RR_l_2        11903 non-null  Float64
4   RR_l_2                11903 non-null  Int64
5   RR_l_2/RR_l_3        11903 non-null  Float64
6   RR_l_3                11903 non-null  Int64
7   RR_l_3/RR_l_4        11903 non-null  Float64
8   RR_l_4                11903 non-null  Int64
9   RR_r_0                11903 non-null  Int64
10  RR_r_0/RR_r_1        11903 non-null  Float64
11  RR_r_1                11903 non-null  Int64
12  RR_r_1/RR_r_2        11903 non-null  Float64
13  RR_r_2                11903 non-null  Int64
14  RR_r_2/RR_r_3        11903 non-null  Float64
15  RR_r_3                11903 non-null  Int64
16  RR_r_3/RR_r_4        11903 non-null  Float64
17  RR_r_4                11903 non-null  Int64
18  seq_size              11903 non-null  Int64
19  signal_mean          11903 non-null  Float64
20  signal_std           11903 non-null  Float64
21  wl_side              11903 non-null  Int64
22  wr_side              11903 non-null  Int64
23  P_val                11903 non-null  Float64
24  Q_val                11903 non-null  Float64
25  R_val                11903 non-null  Float64
26  S_val                11903 non-null  Float64
27  T_val                11903 non-null  Float64
28  P_pos                11903 non-null  Float64
29  Q_pos                11903 non-null  Float64
30  R_pos                11903 non-null  Float64
31  S_pos                11903 non-null  Float64
32  T_pos                11903 non-null  Float64
33  label                11903 non-null  Int64
dtypes: Float64(20), Int64(14)
memory usage: 3.6 MB
```

Pav 2.5. Požymių aprašymas ištrynus eilutes

2.4. Taškai atsiskyrėliai.

Viso ekstremalių atsiskyrėlių: 2190

Bendras ekstremalių atsiskyrėlių procentas:
73.0 %

Atsiskyrėlių procentas kiekvienoje klasėje:
Ekstremalūs atsiskyrėliai 0 klasei: 33.5 %
Ekstremalūs atsiskyrėliai 1 klasei: 87.2 %
Ekstremalūs atsiskyrėliai 2 klasei: 98.3 %

Pav 2.6. Taškų atsiskyrėlių rezultatai.

Paėmus po 1000 įrašų iš kiekvienos klasės, ekstremalių atsiskyrėlių bendras procentas buvo 73% (pav. 2.6). Šį skaičių gerokai padidina 1 ir 2 klasės, kuriuose ekstremalių atsiskyrėlių yra 87,2% ir 98,3% atitinkamai.

Atsiskyrėlių šioje duomenų aibėje nenaikinome, nes duomenys yra jautrūs, t.y. jei reikšmė yra skaitoma kaip atsiskyrėlis, ji nebūtinai yra duomenų triukšmas – tai gali būti širdies veiklos sutrikimas.

2.5. Duomenų aibės normavimas.

Iš kiekvieno label tipo kategorijos atsitiktinai paėmę 1000 reikšmių (pav. 2.6), jas normalizavome. Duomenų normalizavimui pasirinkome šiuos požymius: P_val, Q_val, R_val, S_val, signal_mean ir signal_std reikšmes. Kiti požymiai buvo neliesti, nes jos ir taip pakankamai normalizuotos arba jų normalizavimas nėra logiškas. Duomenys buvo normalizuoti naudojant du metodus:

1. Vidurkio ir dispersijos normalizacija (2.7 pav.);
2. Min-max normalizacija (pav. 2.8).

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	seq_size	signal_mean	signal_std	label
count	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0
mean	0.111981	-0.306704	1.128354	-0.453936	0.372628	0.419863	0.477735	0.531837	212.42	-0.17892	0.324726	1.0
std	0.473079	0.515426	0.616834	0.503136	0.112967	0.105709	0.1068	0.127742	57.815279	0.137566	0.19581	0.816633
min	-1.157755	-2.716884	-0.014881	-2.603065	0.055	0.04	0.06	0.125	100.0	-1.230984	0.069568	0.0
25%	-0.202918	-0.455469	0.632942	-0.589376	0.31	0.355	0.425	0.49	183.0	-0.24754	0.158867	0.0
50%	-0.095403	-0.255238	0.996104	-0.46382	0.35	0.455	0.5	0.53	205.0	-0.178909	0.264662	1.0
75%	0.392424	-0.133059	1.580117	-0.263665	0.445	0.475	0.515	0.58	230.0	-0.108514	0.467014	2.0
max	2.496907	2.658497	2.9692	2.69908	0.86	0.845	0.915	0.985	928.0	1.351685	1.070534	2.0

Pav 2.7. Tyriamųjų duomenų aprašomoji statistika

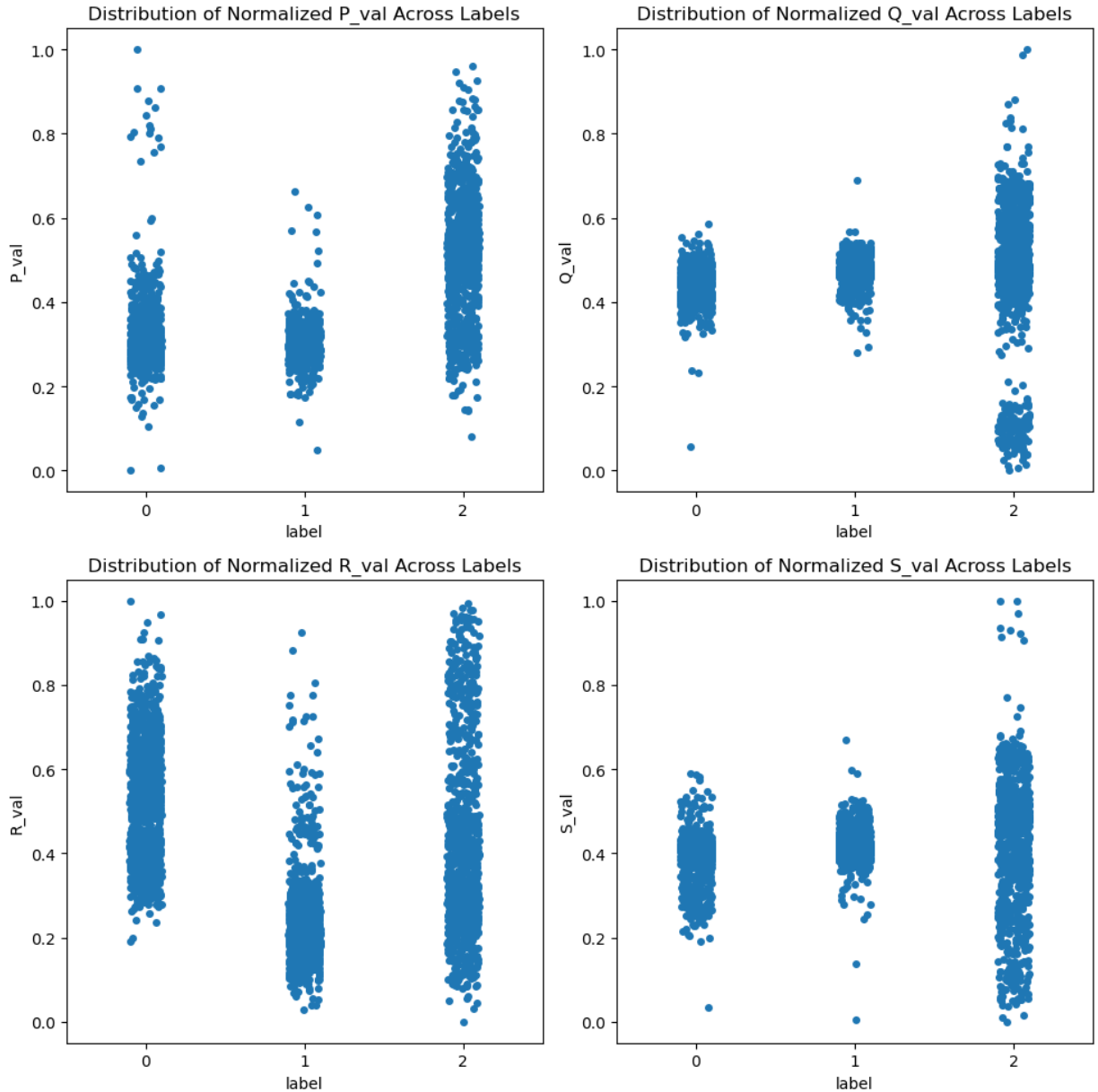
	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	seq_size	signal_mean	signal_std	label
count	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0
mean	0.0	0.0	-0.0	-0.0	0.372628	0.419863	0.477735	0.531837	212.42	0.0	0.0	1.0
std	1.0	1.0	1.0	1.0	0.112967	0.105709	0.1068	0.127742	57.815279	1.0	1.0	0.816633
min	-2.683984	-4.676093	-1.853391	-4.27147	0.055	0.04	0.06	0.125	100.0	-7.647723	-1.303091	0.0
25%	-0.665639	-0.288626	-0.803153	-0.269192	0.31	0.355	0.425	0.49	183.0	-0.498818	-0.847043	0.0
50%	-0.438373	0.099851	-0.214401	-0.019644	0.35	0.455	0.5	0.53	205.0	0.00008	-0.306745	1.0
75%	0.592802	0.336895	0.73239	0.378171	0.445	0.475	0.515	0.58	230.0	0.5118	0.726658	2.0
max	5.041283	5.752913	2.984346	6.266734	0.86	0.845	0.915	0.985	928.0	11.126354	3.808825	2.0

Pav 2.8. Z-score normalizuotų aprašomoji statistika

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	seq_size	signal_mean	signal_std	label
count	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0	3000.0
mean	0.347429	0.448374	0.383111	0.405332	0.372628	0.419863	0.477735	0.531837	212.42	0.407355	0.254912	1.0
std	0.129445	0.095886	0.206708	0.094893	0.112967	0.105709	0.1068	0.127742	57.815279	0.053265	0.195621	0.816633
min	0.0	0.0	0.0	0.0	0.055	0.04	0.06	0.125	100.0	0.0	0.0	0.0
25%	0.261265	0.420699	0.217093	0.379788	0.31	0.355	0.425	0.49	183.0	0.380786	0.089213	0.0
50%	0.290684	0.457948	0.338793	0.403468	0.35	0.455	0.5	0.53	205.0	0.40736	0.194907	1.0
75%	0.424165	0.480677	0.534502	0.441218	0.445	0.475	0.515	0.58	230.0	0.434616	0.397062	2.0
max	1.0	1.0	1.0	1.0	0.86	0.845	0.915	0.985	928.0	1.0	1.0	2.0

Pav 2.9. min-max normalizuotų duomenų aprašomoji statistika

2.6. Duomenų aibės vizualizacija, analizė, išvados.

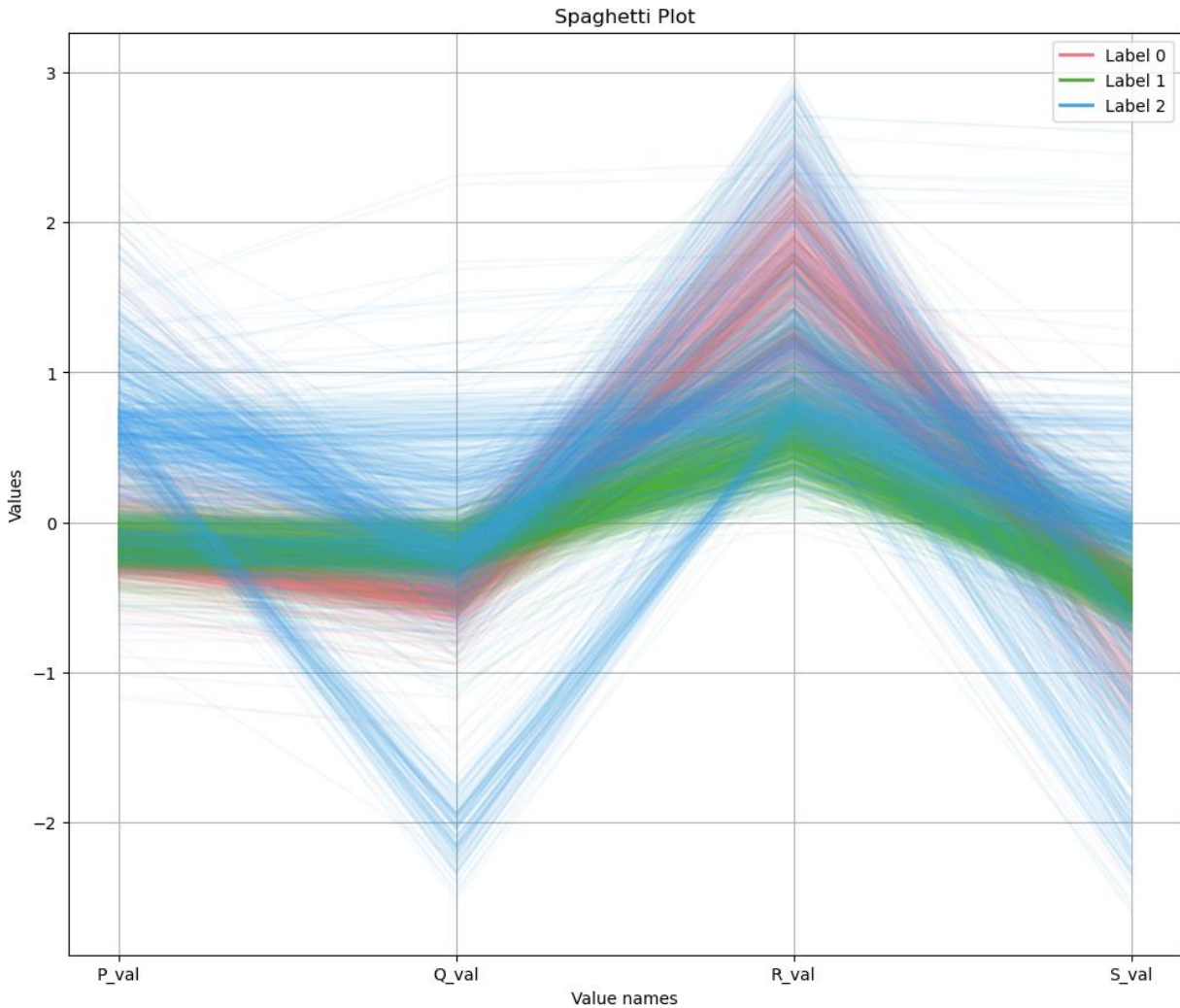


Pav 2.10. PQRS verčių paskalidytų taškų diagrama

Remdamiesi pateiktu keturių diagramų rinkiniu (pav. 2.10), galime padaryti keletą išvadų apie P, Q, R ir S bangų pasiskirstymą pagal skirtingas „label“ kategorijas (0, 1, 2):

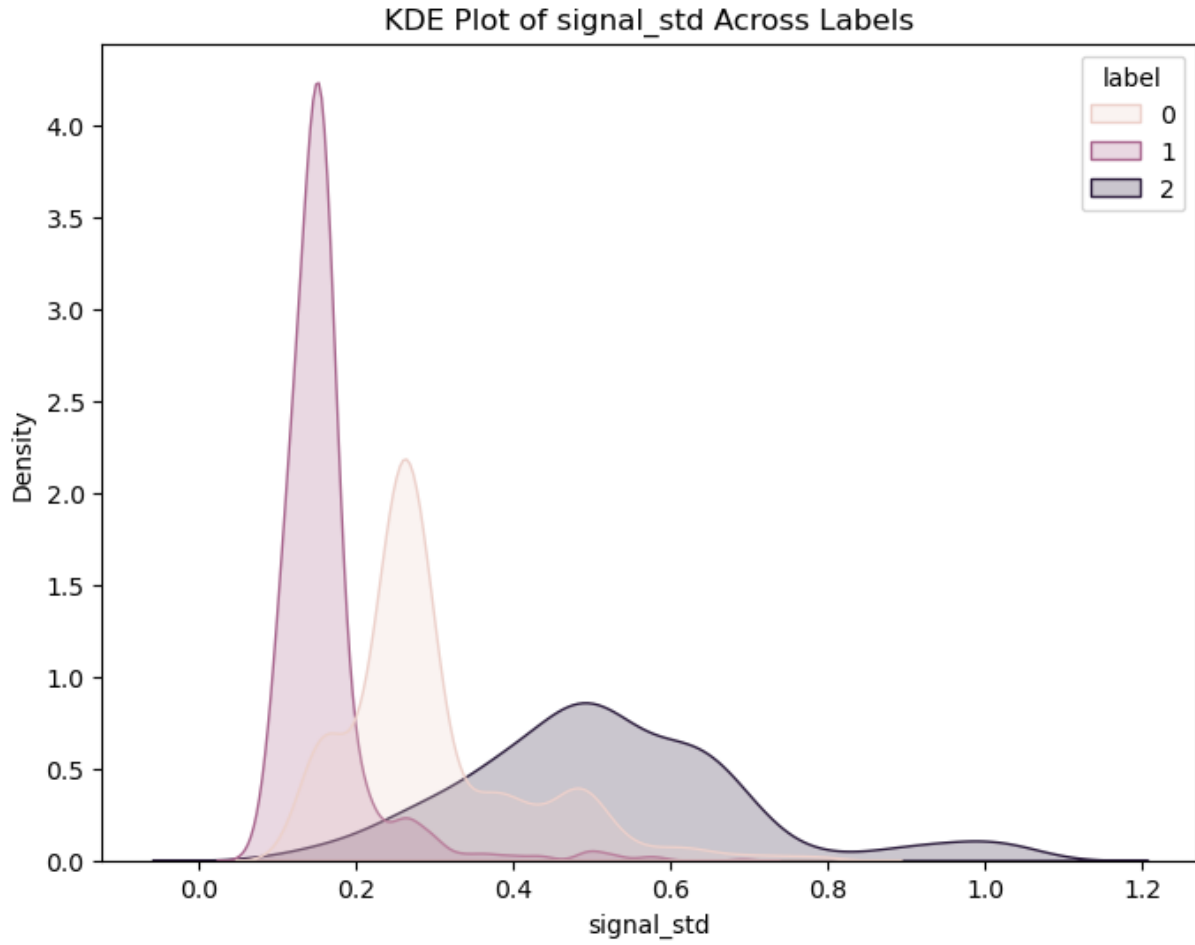
1. P_val pasiskirstymas: P_val pasiskirstymas „label 2“ kategorijoje pasireiškia didžiausiomis variacijomis – didžioji dalis normalizuotų reikšmių priklauso intervalui $[0,2; 0,8]$.
2. Q_val pasiskirstymas: Q_val pasiskirstymas „label 2“ kategorijoje pasireiškia didžiausiomis variacijomis – $[0; 0,2]$ ir $[0,3; 0,7]$, o „label 0“ ir „label 1“ beveik visos reikšmės priklauso $[0,3; 0,55]$.

3. R_val pasiskirstymas: R_val pasiskirstymas visose kategorijose yra panašus, tačiau „label 0“ kategorijoje reikšmės nenukrenta žemiau 0,2.
4. S_val pasiskirstymas: S_val pasiskirstymas „label 2“ kategorijoje pasireiškia didžiausiomis variacijomis – didžioji dalis reikšmių patenka į intervalą [0; 0,7], o „label 0“ ir „label 1“ kategorijose visos reikšmės linkusios koncentruotis intervale [0,2; 0,6].

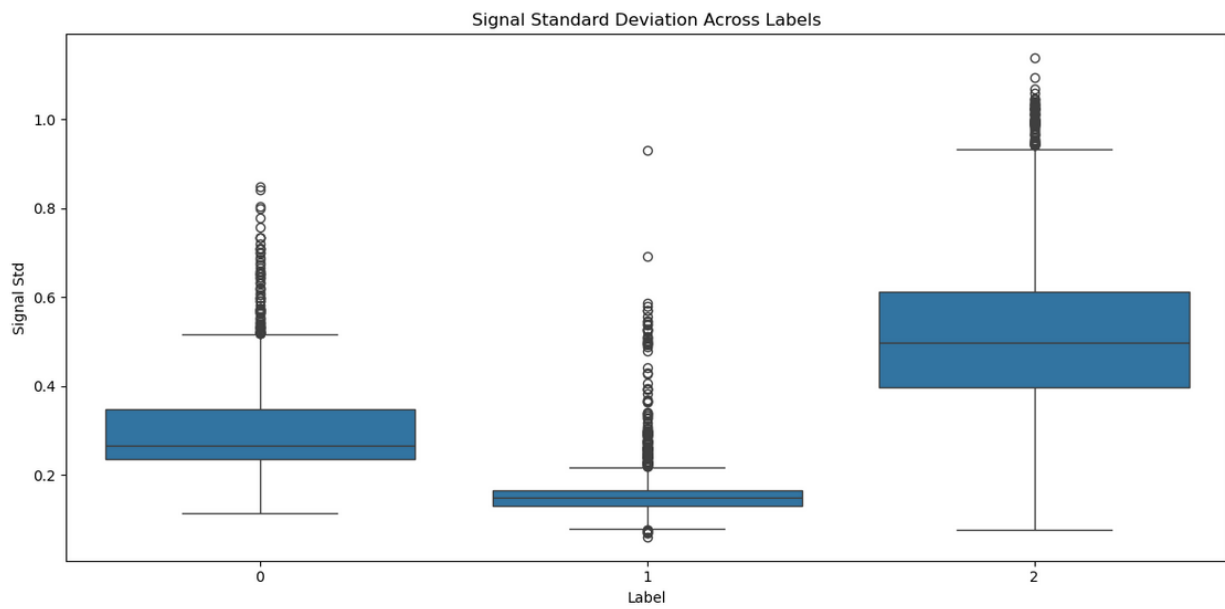


Pav 2.10.11. PQRS verčių paskalidytų taškų diagrama

Panašias išvadas matome ir šioje spagečių diagramoje (pav. 2.11): „label 0“ ir „label 1“ pasižymi stabilėniu ir mažiau įvairiu reikšmių pasiskirstymu, o „label 2“ reikšmės yra pasiskirsčiusios plačiau, ypač ties Q ir S bangomis.



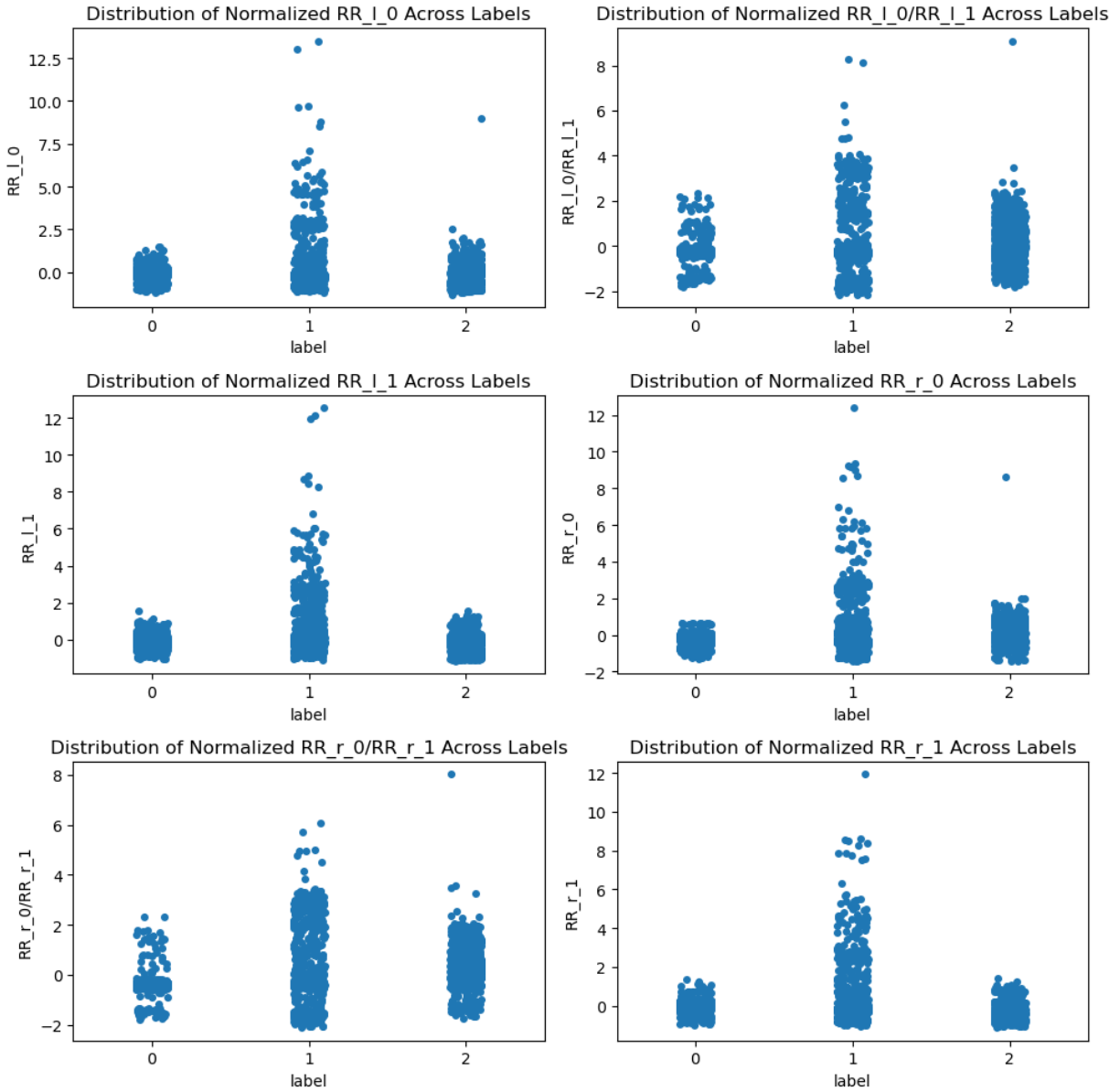
Pav 2.12. signal_std branduolinio tankio diagrama



Pav 2.13. Stačiakampė diagrama

1. „Label 0“ kategorijos signal_std reikšmė branduolinio tankio diagramoje (pav. 2.12) yra išsidėsčiusi platesniame diapazone – intervale $[0,1; 0,5]$, kas reiškia, kad ji turi gan didelę variaciją. Tas pats matosi ir stačiakampėje diagramoje (pav. 2.13) – 0,25 ir 0,75 kvartilės „Label 0“ kategorijoje yra intervale $[0,1; 0,5]$.
2. „Label 1“ kategorijoje reikšmės yra stabilesnės – jos pasiskirsčiusios intervale $[0,1; 0,2]$.
3. „Label 2“ kategorija išsiskiria daug platesniu reikšmių pasiskirstymu, su didesniu piku intervale $[0,4; 0,6]$ ir ilgesne uodega, kuri tęsiasi iki 1.1. Tai reiškia, kad „label 2“ kategorija turi didelę standartinio nuokrypio variaciją.

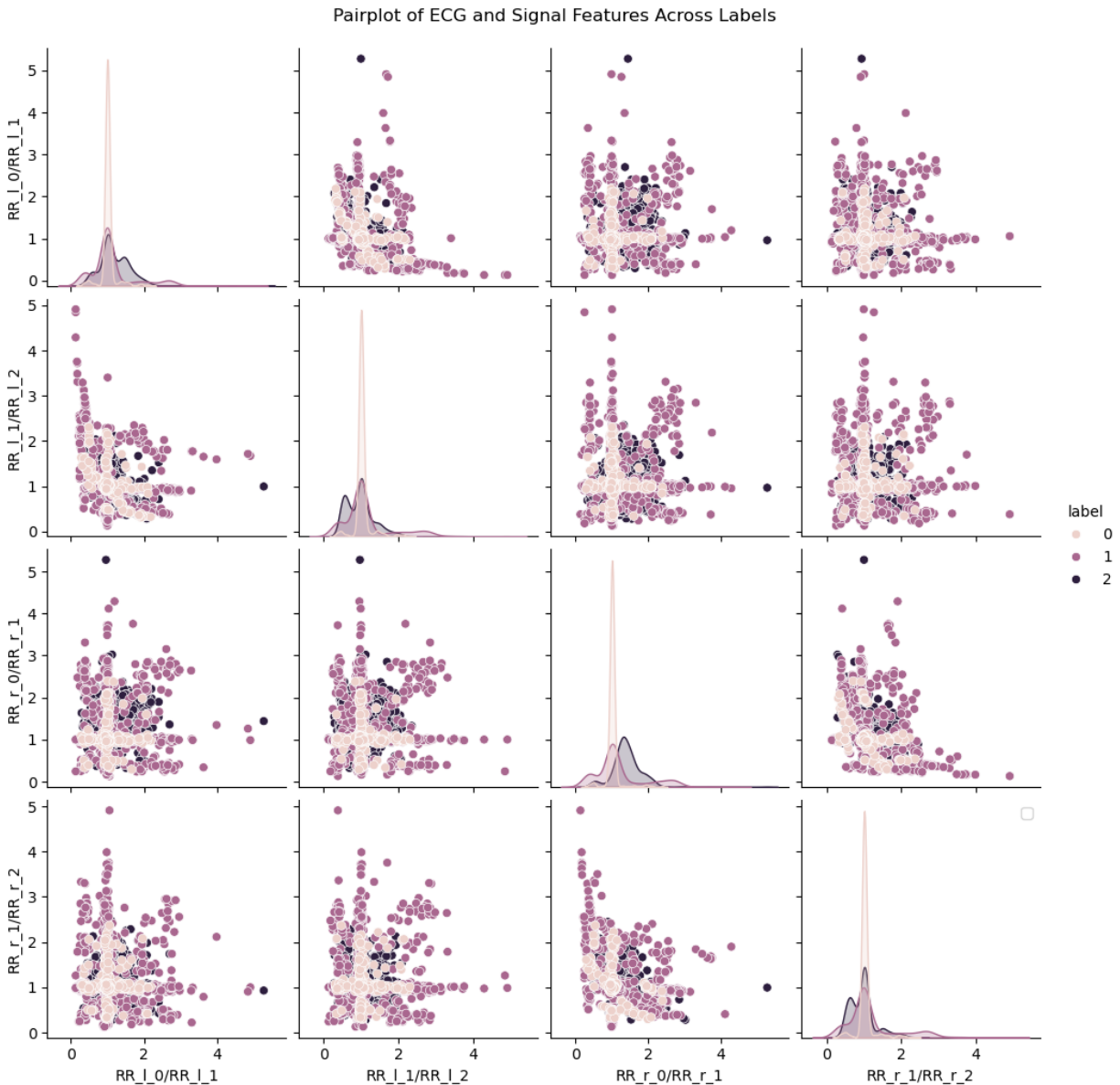
Matome, kad „label 1“ pūpsnių signalų standartiniai nuokrypiai absoliučiai didžiąją dalį laiko bus žemiau 0,2. Didėjančios standartinių nuokrypių reikšmės reikš didesnę galimybę pūpsniui patekti į „label 0“ kategoriją, o perkopus ją – į „label 2“.



Pav 2.14. pasklaidytų taškų diagrama

Sekančioje, pasklaidytų taškų diagramoje (Pav. 2.14), „Label 1“ kategorija aiškiai išsiskiria visuose rodikliuose su didžiausia variacija ir reikšmingai aukštesnėmis RR intervalų ir jų santykių vertėmis, lyginant su „Label 0“ ir „Label 2“. Tai reiškia, kad šiai grupei būdingos didesnės R bangų intervalo variacijos.

„Label 0“ ir „Label 2“ kategorijos turi daug mažesnę kintamumą, vertės turi mažesnę variaciją ir reikšmę.



Pav 2.15. RR santykių porinė sklaidos ir branduolinio tankio diagrama

RR santykių porinės sklaidos ir branduolinio tankio diagramų (Pav. 2.15) išvados:

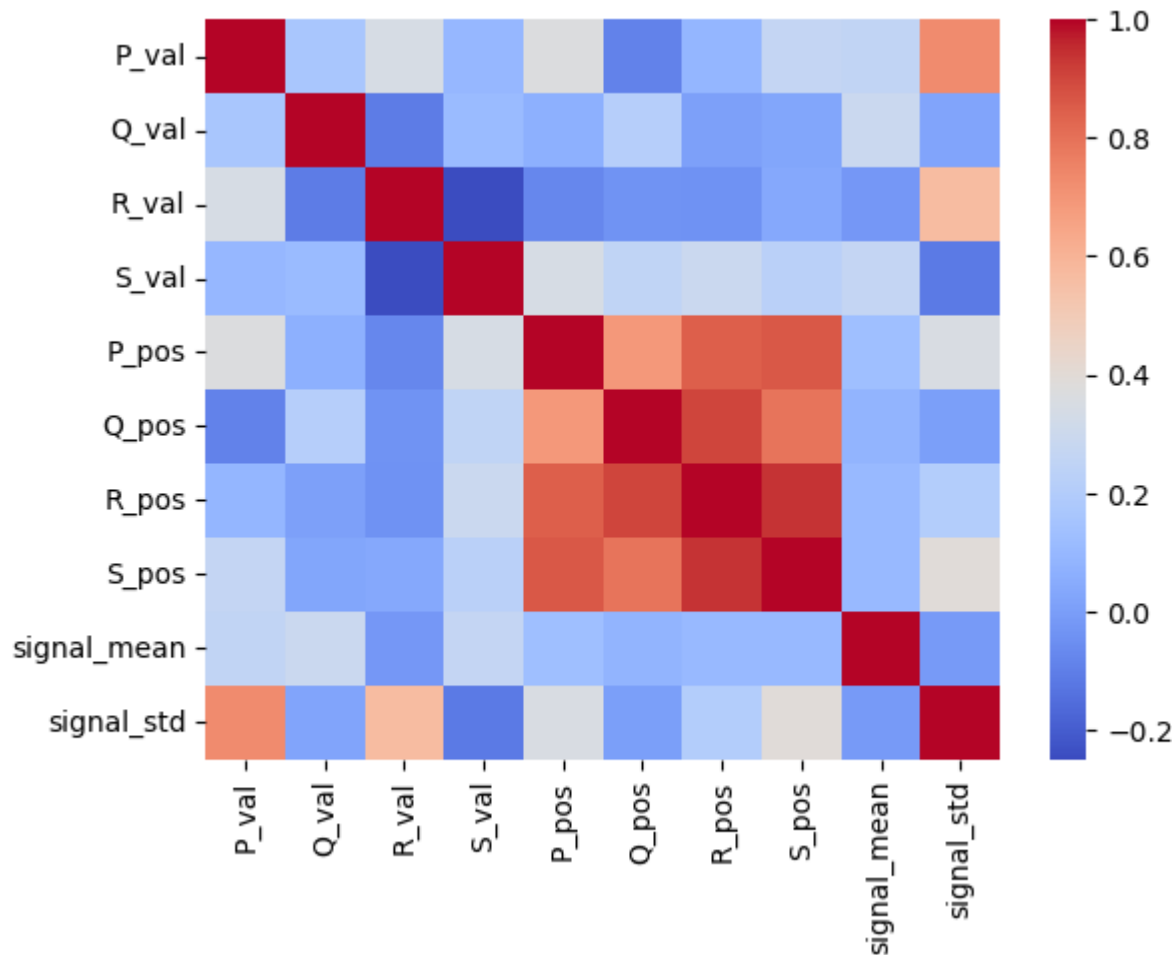
1. „Label 0“ (šviesi spalva) kategorijos duomenys turi tendenciją pasiskirstyti arčiau vieno RR intervalų santykių srityje branduolinio tankio diagramoje (toliau BTD). Dauguma „label 0“ taškų sutelkti gana arti centro, o jų pasiskirstymas yra gana vienodas porinės sklaidos diagramoje.
2. „Label 1“ (šviesiai violetinė spalva) ir „label 2“ (tamsiai violetinė spalva) kategorijos duomenys turi daug platesnį pasiskirstymą BTD. Didžioji dalis šių kategorijų santykių reikšmių pasiskirsčiusios intervale [0; 1,5], tačiau tęsiasi iki 3.

3. „Label 1“ kategorija porinės sklaidos diagramoje turi didžiausią pasiskirstymą, tai rodo didelę variaciją. Ši kategorija apima daugiausia taškų, kurie yra atitolę nuo nulio, ypač ties RR santykiais, kurie yra virš 2.

2.7. Požymių tarpusavio priklausomybės analizė.

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	signal_mean	signal_std
P_val	1.000000	0.080527	0.346926	0.074815	0.360962	-0.107093	0.084010	0.253416	0.227289	0.720938
Q_val	0.080527	1.000000	-0.096394	0.053806	-0.014868	0.198313	-0.055960	-0.053947	0.236791	-0.058059
R_val	0.346926	-0.096394	1.000000	-0.269891	-0.098748	-0.046898	-0.061700	0.012613	-0.038582	0.560590
S_val	0.074815	0.053806	-0.269891	1.000000	0.352741	0.224242	0.287011	0.222113	0.296790	-0.112685
P_pos	0.360962	-0.014868	-0.098748	0.352741	1.000000	0.654663	0.838425	0.856461	0.100584	0.354670
Q_pos	-0.107093	0.198313	-0.046898	0.224242	0.654663	1.000000	0.892243	0.768307	0.040816	-0.017890
R_pos	0.084010	-0.055960	-0.061700	0.287011	0.838425	0.892243	1.000000	0.937585	0.064407	0.193323
S_pos	0.253416	-0.053947	0.012613	0.222113	0.856461	0.768307	0.937585	1.000000	0.043172	0.387693
signal_mean	0.227289	0.236791	-0.038582	0.296790	0.100584	0.040816	0.064407	0.043172	1.000000	-0.083503
signal_std	0.720938	-0.058059	0.560590	-0.112685	0.354670	-0.017890	0.193323	0.387693	-0.083503	1.000000

Pav 2.16. Koreliacijos tarp reikšmių lentelė, visos „Label“ kategorijos



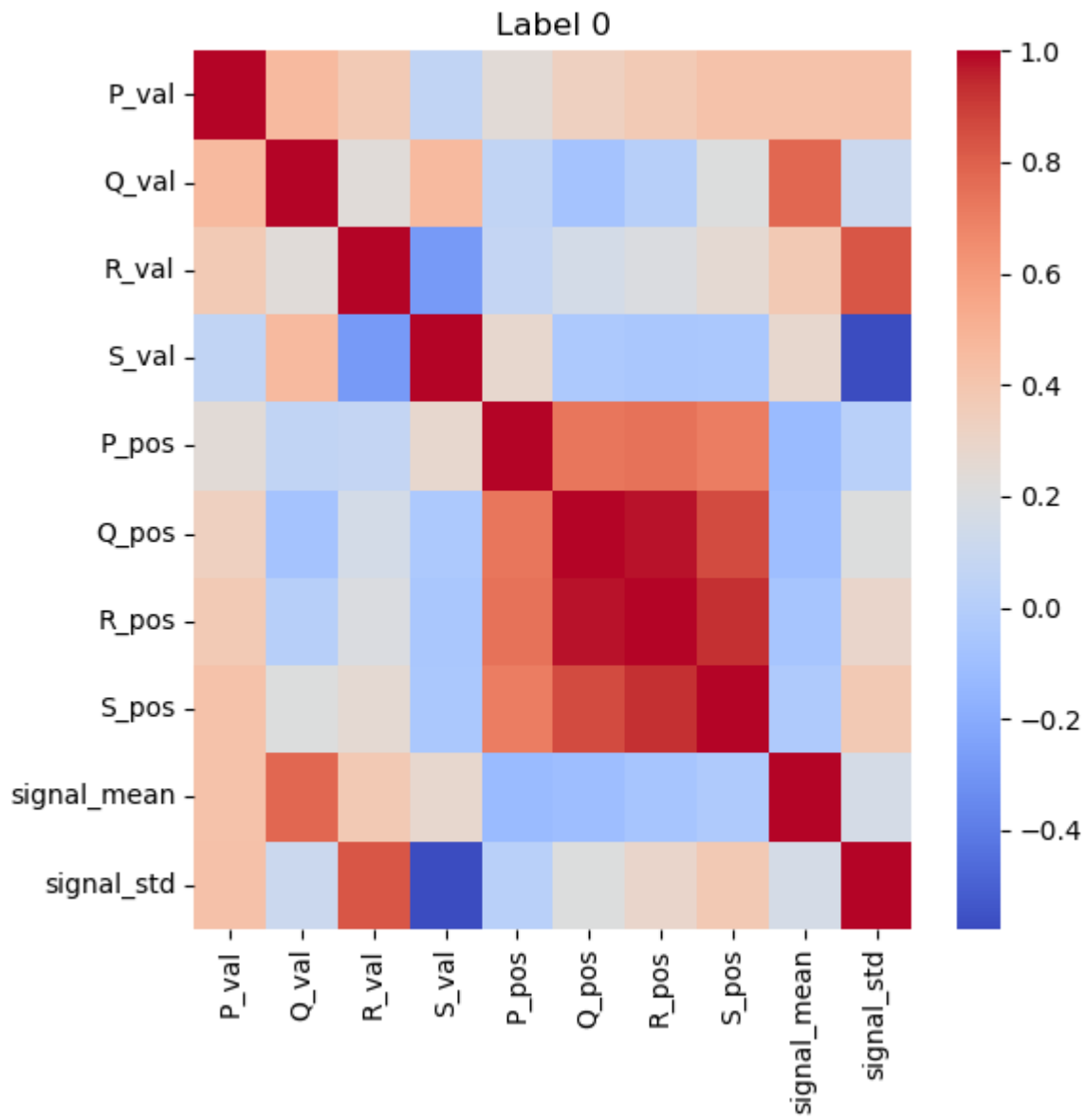
Pav 2.17. Koreliacija tarp reikšmių, visos „Label“ kategorijos

Pirmas korelacių grafikas rodo bendras atrinktų reikšmių korelacijas, neatsižvelgiant į „label“ reikšmę (pav. 2.16, pav. 2.17).

Pastebimos ryškiausios korelacijos tarp R_pos ir S_pos – 0,93 , taip pat stipri korelacija tarp R_pos ir Q_pos – 0,89. Apibendrinant visas pozicijų reikšmes, galime tvirtai teigti, kad jų tarpusavio korelacija yra stipri – padidėjus vienai, didėja ir kita. Taip pat pakankamai stipri korelacija yra tarp P_val ir signal_std – siekianti 0,72.

Mėlyna spalva rodo neigiamas korelacijas, bet labai nedideles. Didžiausia jų yra tarp signal_std ir P_val (-0,26).

Likusios reikšmės ypač reikšmingų rezultatų nesuteikia, t.y. didelės korelacijos neturi.

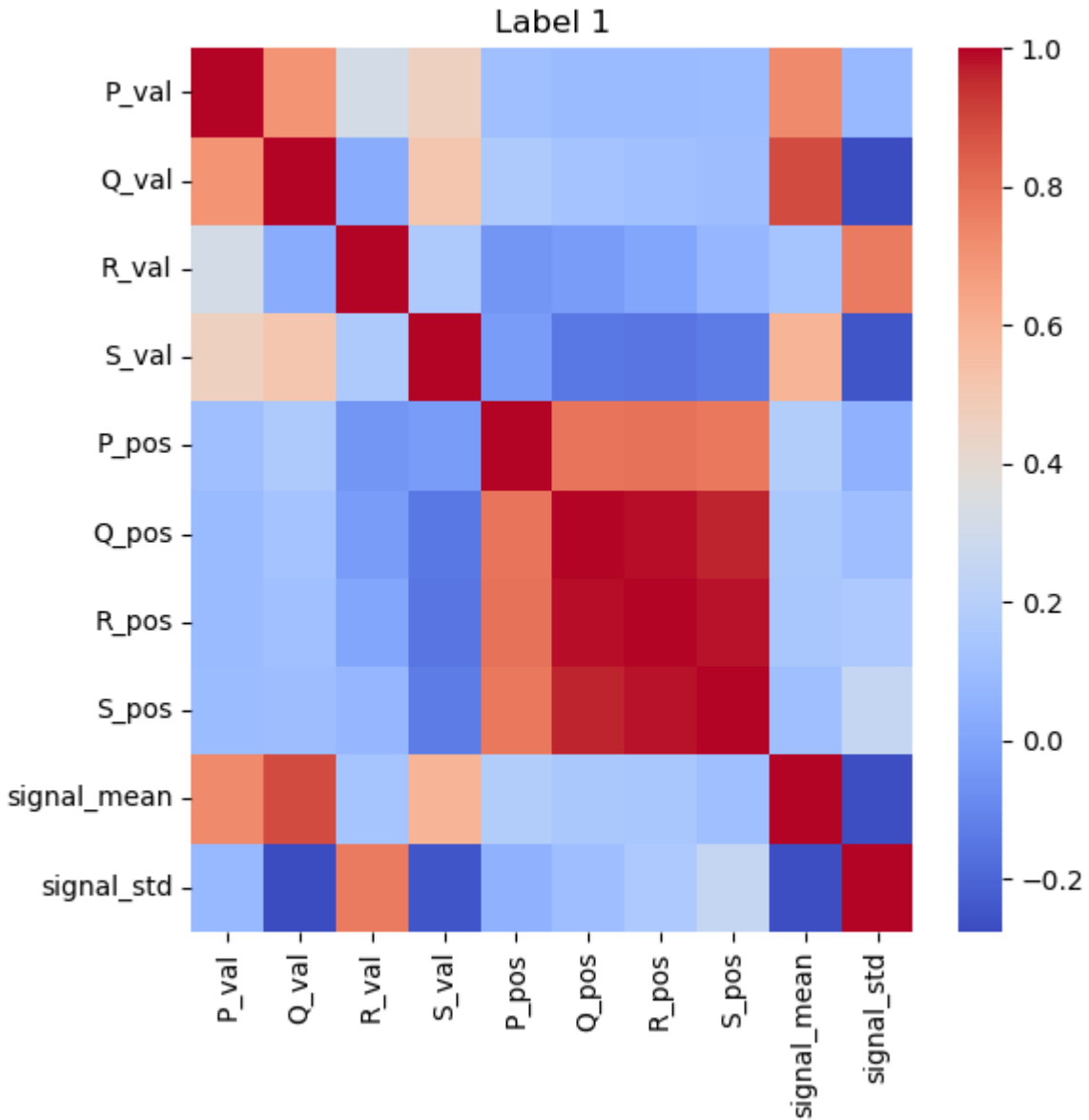


Pav 2.18. "Label 0" kategorijos koreliacija

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	signal_mean	signal_std
P_val	1.000000	0.460414	0.372060	0.060617	0.242245	0.334022	0.375981	0.418164	0.418223	0.423573
Q_val	0.460414	1.000000	0.231620	0.463052	0.063545	-0.074768	0.010998	0.207400	0.782069	0.106778
R_val	0.372060	0.231620	1.000000	-0.278096	0.072937	0.150344	0.197585	0.256587	0.382209	0.829530
S_val	0.060617	0.463052	-0.278096	1.000000	0.275598	-0.040574	-0.052396	-0.043260	0.274083	-0.577008
P_pos	0.242245	0.063545	0.072937	0.275598	1.000000	0.732228	0.744826	0.706578	-0.125987	0.014619
Q_pos	0.334022	-0.074768	0.150344	-0.040574	0.732228	1.000000	0.980874	0.863926	-0.108107	0.205421
R_pos	0.375981	0.010998	0.197585	-0.052396	0.744826	0.980874	1.000000	0.926926	-0.064946	0.287859
S_pos	0.418164	0.207400	0.256587	-0.043260	0.706578	0.863926	0.926926	1.000000	-0.023332	0.379068
signal_mean	0.418223	0.782069	0.382209	0.274083	-0.125987	-0.108107	-0.064946	-0.023332	1.000000	0.157419
signal_std	0.423573	0.106778	0.829530	-0.577008	0.014619	0.205421	0.287859	0.379068	0.157419	1.000000

Pav 2.19. "Label 0" kategorijos koreliacija lentelės pavidalu

Sekančiose lentelėse (pav. 2.18, pav. 2.19) matome panašias koreliacijas. Pozicijų tarpusavio koreliacijos yra šiek tiek stipresnės už visų „label“ kategorijų koreliacijas. Signal_std ir s_val reikšmės turi didžiausią neigiamą koreliaciją iš visų kategorijų (-0,58).

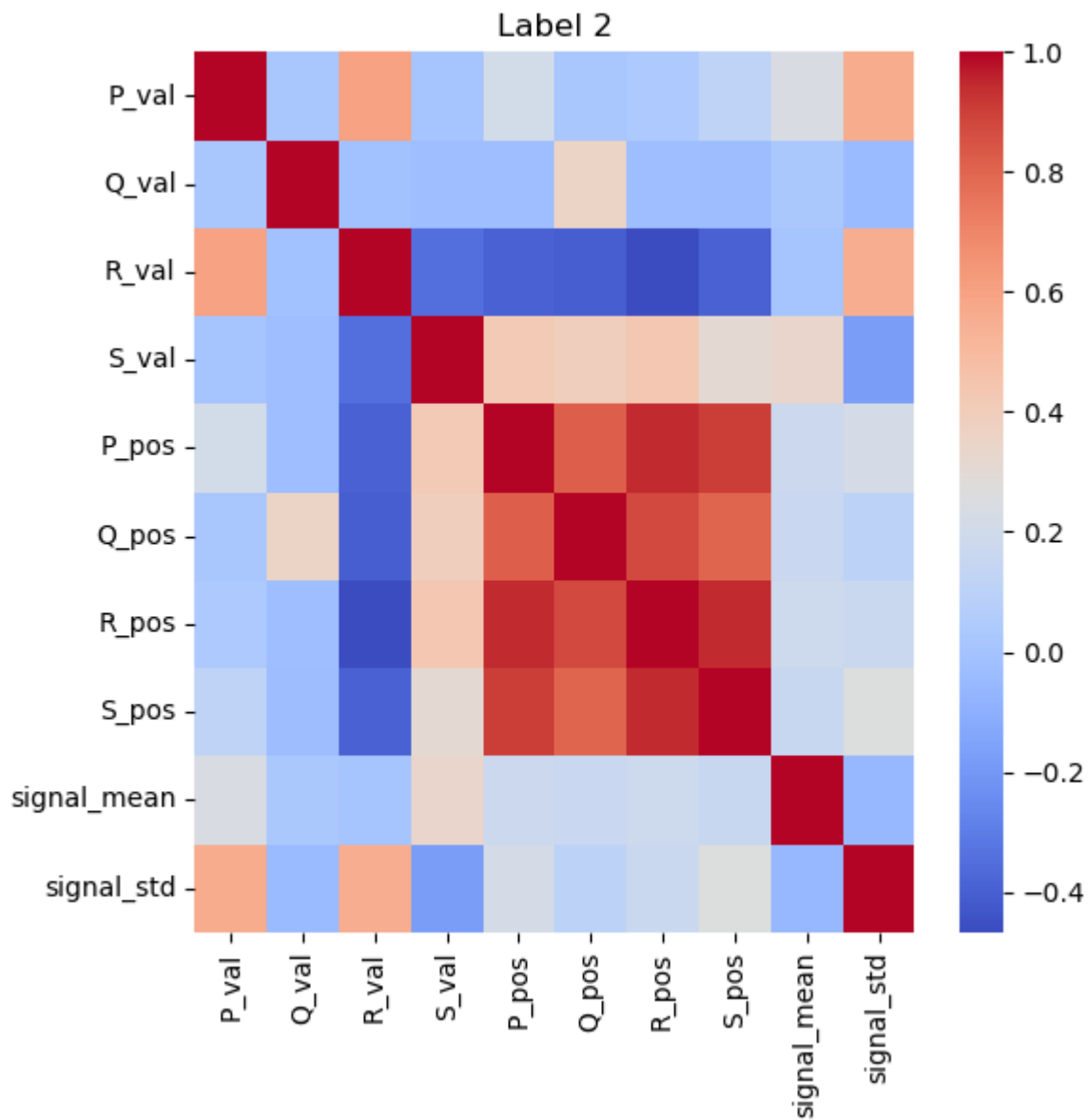


Pav 2.20. "Label 1" kategorijos koreliacija

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	signal_mean	signal_std
P_val	1.000000	0.699185	0.314721	0.457529	0.111174	0.092270	0.091773	0.096246	0.729825	0.084295
Q_val	0.699185	1.000000	0.031391	0.512011	0.165783	0.128586	0.115205	0.098258	0.889293	-0.275599
R_val	0.314721	0.031391	1.000000	0.164092	-0.046440	-0.031359	0.008343	0.077745	0.135950	0.767332
S_val	0.457529	0.512011	0.164092	1.000000	-0.030638	-0.145676	-0.152915	-0.133575	0.589037	-0.248727
P_pos	0.111174	0.165783	-0.046440	-0.030638	1.000000	0.789448	0.792546	0.780541	0.184923	0.050627
Q_pos	0.092270	0.128586	-0.031359	-0.145676	0.789448	1.000000	0.988580	0.964872	0.157111	0.105433
R_pos	0.091773	0.115205	0.008343	-0.152915	0.792546	0.988580	1.000000	0.982942	0.149599	0.158907
S_pos	0.096246	0.098258	0.077745	-0.133575	0.780541	0.964872	0.982942	1.000000	0.110633	0.256684
signal_mean	0.729825	0.889293	0.135950	0.589037	0.184923	0.157111	0.149599	0.110633	1.000000	-0.270614
signal_std	0.084295	-0.275599	0.767332	-0.248727	0.050627	0.105433	0.158907	0.256684	-0.270614	1.000000

Pav 2.21. „Label 1” kategorijos koreliacija lentelės pavidalu

„Label 1“ korelacių diagramose (pav. 2.20, pav. 2.21) matomos išskirtinai stiprios Q_pos, R_pos, S_pos tarpusavio koreliacijos – jos yra beveik tiesioginės (>0,96). Koreliacijos tarp P_pos ir likusių pozicijų yra panašios į „Label 0“ grafike esančias koreliacijas (maždaug 0,8).



Pav 2.22. "Label 2" kategorijos koreliacija

	P_val	Q_val	R_val	S_val	P_pos	Q_pos	R_pos	S_pos	signal_mean	signal_std
P_val	1.000000	0.019155	0.601903	0.007349	0.204494	0.015968	0.035313	0.114899	0.239737	0.561026
Q_val	0.019155	1.000000	-0.013369	-0.030641	-0.025857	0.353994	-0.026010	-0.033253	0.026968	-0.048598
R_val	0.601903	-0.013369	1.000000	-0.350074	-0.394107	-0.407341	-0.467121	-0.393125	0.007181	0.555383
S_val	0.007349	-0.030641	-0.350074	1.000000	0.418300	0.391974	0.432649	0.311776	0.340600	-0.176615
P_pos	0.204494	-0.025857	-0.394107	0.418300	1.000000	0.819075	0.944470	0.904169	0.174056	0.217078
Q_pos	0.015968	0.353994	-0.407341	0.391974	0.819075	1.000000	0.876229	0.804071	0.158692	0.104407
R_pos	0.035313	-0.026010	-0.467121	0.432649	0.944470	0.876229	1.000000	0.945122	0.181477	0.164228
S_pos	0.114899	-0.033253	-0.393125	0.311776	0.904169	0.804071	0.945122	1.000000	0.154571	0.266071
signal_mean	0.239737	0.026968	0.007181	0.340600	0.174056	0.158692	0.181477	0.154571	1.000000	-0.055876
signal_std	0.561026	-0.048598	0.555383	-0.176615	0.217078	0.104407	0.164228	0.266071	-0.055876	1.000000

Pav 2.23. „Label 2” kategorijos koreliacija lentelės pavidalu

„Label 2” grafikuose (pav. 2.22, pav. 2.23) išsiskiria S_val, P_pos, R_pos, Q_pos, S_pos neigiamos koreliacijos su R_val – čia jos svyruoja intervale [-0,46; -0,35].

Taip pat galime pastebėti stiprias teigiamas koreliacijas tarp P_pos, Q_pos, R_pos ir S_pos verčių.

Bendros visų koreliacijų išvados: neigiamos s_val koreliacijos su signal_std, ypač „Label 0“ kategorijoje (-0,58). Stiprios teigiamos Q_pos, R_pos, S_pos tarpusavio koreliacijos, ypač „Label 1” kategorijoje (>0,96). S_pos, P_pos, R_pos, Q_pos neigiamos koreliacijos su S_val [-0,15; -0,03], o „Label 2” kategorijoje teigiamos koreliacijos (~0.4). „Label 2” kategorijoje S_val, P_pos, R_pos, Q_pos, S_pos neigiamos koreliacijos su R_val [-0,46; -0,35].

3. IŠVADOS

- „label 2” turi didžiausią P, Q ir S verčių pasiskirstymą, taip pat atskirų įrašų P, Q, R ir S reikšmių skirtumai yra nutolę nuo vidurkio lyginant su „label 0“ ir „label 2”, t.y. reikšmės arba labai panašios, arba ypač skiriasi. Taip pat „label 2” signalo standartinio nuokrypių reikšmės turi labai didelę variaciją – intervalas siekia [0,1; 1,1] ir sąlyginai aukštą vidurkį. Tačiau RR verčių santykių variacija yra panaši į „label 0“. Taip pat galime pastebėti stipresnę nei vidutinę pozicijų ir S vertės neigiamą koreliaciją su R_val, ko visai nėra kitose kategorijose.
- „label 1” P, Q, R ir S vertės neišsiskiria, jos varijuoja panašiai, kaip ir „label 0“, tačiau spagečių diagramoje (pav. 2.11) galime pastebėti, kad atskirų įrašų P, Q, R ir S reikšmių skirtumai mažesni lyginant su „label 0“ ir „label 2“. Taip pat labai aiškiai matome signal_std branduolinio tankio diagramoje (pav. 2.12), kad „label 1” didžioji dalis reikšmių patenka į intervalą [0,1; 0,2], jų vidurkis yra pats mažiausias (0,16). RR intervalų ir jų santykių variacija taip pat labai išsiskiria iš kitų kategorijų, turi įvairesnę ir platesnę reikšmių verčių pasiskirstymą. Taip pat išsiskiria Q_pos, R_pos ir S_pos tarpusavio koreliacijos – jos yra beveik tiesioginės (>0,96).
- Tuo tarpu kategorijoje „label 0“ labai didelių išskirčių nepastebėta, jų reikšmių variacija dažniausiai yra mažiausia ir jų vidurkiai linkę būti vidutiniai. Taip pat RR intervalų santykių branduolinėje tankio diagramoje (pav. 2.15) aiškiai matome, kad santykių variacija yra mažiausia ir beveik visų santykių vidurkis yra intervale [1,1; 1,2], tuo tarpu kitų kategorijų variacija didelė ir didžioji dalis reikšmių patenka į intervalą [0; 2].