

Bangladeshi naiste tervisenäitajate mõju rasedusaegsele riskihinnangule

Projekti tiim: Hannes Aaresild, Natalja Beljajeva, Mari Helen Štarkov

GitHubi link: <https://github.com/aaresild/maternal-health-risk#maternal-health-risk>

1. Business Understanding

1.1 Taust

Uurime projektis Bangladeshi rasedate naiste tervisenäitajaid ja erinevate tervisenäitajate mõju rasedusaegsele riskihinnangule. Andmed on kogutud Bangladeshi maapiirkondade haiglatest, kliinikutest ja emaduskeskustest läbi IoT riski monitoorimise süsteemi. Tervisenäitajate mõõtmised võivad toimuda ka kodustes tingimustes.

Bangladeshis on puudus oskuslikest sünnitusabistajatest, eriti maapiirkondades. Bangladesh kaotab umbes 7 660 naist igal aastal ennetatavate põhjuste tõttu, mis on seotud raseduse ja sünnitusega. Bangladeshi emade suremuse määr on 245 naist 100 000 elussünni kohta. (The Lancet 2015) Bangladeshil on erakordselt kõrge noorukite sünnituse määr: 11,3% naistest sünnitab oma esimese lapse vanuses 15–19 aastat. (Girls Not Brides 2017)

Umbes 71% sünnitustest Bangladeshis toimub kodus, millest vaid 4% juhtudest on kohal väljaõppe saanud spetsialist. Maapiirkonnas on olukord veelgi trastilisem, kodusünnituste osakaal moodustab ligi 90%. (Bangladesh DHS 2014) Rohkem kui pooled naised ei pääse raseduse ajal tervishoiuasutustesse, vaid 11% emadest kasutavad kuue nädala jooksul pärast sünnitust tervishoiuteenuseid (HOPE Foundation).

Allikas:

https://everymothercounts.org/grants/bangladesh-a-deeper-dive/?fbclid=IwAR0u9NEdMFeHH7YgG_rdjIJ3ogMvCJangMJGfTAu3zpsbCv20aIGfKS9BP0#:~:text=In%20Bangladesh%2C%20the%20rate%20of%20maternal%20mortality%20is,from%20preventable%20causes%20related%20to%20pregnancy%20and%20childbirth

1.2 Eesmärk

Eesmärgiks on leida seos andmestiku tunnuste ja Bangladeshi naiste rasedusaegse riskihinnangu vahel. Soovime leida vastused järgmistele uurimisküsimustele:

1. Kuidas mõjutavad andmestikus olevad tervisetunnused riskihinnangut?

2. Millised tunnused mõjutavad riskihinnangut olulisemalt?
3. Millised tunnused koos mõjutavad enam riskihinnangut?

1.3 Projekti edu kriteerium

Lõpp-tulem võiks anda hinnangu missuguste tunnuste muutmisel on võimalik saada:

1. Kõrge riski klassist keskmise riski klassi.
2. Keskmise riski klassist madala riski klassi.

Lisaks soovime luua mudeli meditsiinitöötajatele, mis annab naise tervisetunnuseid sisestades tulemuseks riskihinnangu.

Selle info põhjal saab Bangladeshis valitsus võimaluse aidata kindlate riskirühmade esindajaid.

1.4 Andmestik

Andmestikuna kasutame Kaggle keskkonnast Maternal Health Risk andmeid (<https://www.kaggle.com/datasets/joebeachcapital/maternal-health-risk/>).

1.5 Nõuded, eeldused, piirangud

Me pole päris kindlad valimi piisavuses – kas 1014 naise andmed on esinduslikud, et teha põhjanevaid järeldusi. Samuti, kas tunnuste arv ja valik on piisavad, et nende põhjal teha põhjanevaid järeldusi.

Mitte meditsiini alal töötavate inimestena võib olla meil keeruline hinnata, kas tekkinud seosed on päriselt põhjuslikud või juhuslikud.

1.6 Riskid

Valimi väiksuse tõttu võib esineda hinnangutes ebatäpsuseid. Selle teadvustamiseks arvutame analüüsi osas eksimismäära.

Meil, kui mitte meedikutel võib olla keeruline hinnata, kas tunnus korreleerub riskiteguriga või on põhjuslikus seoses.

1.7 Terminoloogia

- *Age* – ema vanus
- *SystolicBP* - Süstoolne vererõhk näitab arterites olevat rõhku siis, kui süda tõmbub kokku ja pumpab verd arteritesse.
- *DiastolicBP* - Diastoolne vererõhk näitab arterites olevat rõhku siis, kui süda on lõõgastunud ja täitub verest.

Mõõtmine mmHg (millimeetrit elavhõbedasammast)

Optimaalne: alla 120/80 mmHg,

Normaalne: 120-129/80-84 mmHg,

Kõrge normaalne: 130-139/85-89 mmHg,

1. astme hüpertensioon: 140-159/90-99 mmHg,

2. astme hüpertensioon: 160-179/100-109 mmHg,

3. astme hüpertensioon: 180 või kõrgem/110 või kõrgem mmHg,

Vahe süstoolse ja diastoolse rõhu vahel nimetatakse pulsisurevks.

Tavaliselt on pulsisure vahemikus 30 kuni 50 mmHg. See tähendab, et kui inimese süstoolne rõhk on 120 mmHg ja diastoolne rõhk on 80 mmHg, siis pulsisure on 40 mmHg.

Pulsisure, mis on liiga kõrge (üle 60 mmHg), on seotud suurenenud riskiga südamehaiguste, insuldi ja suremuse osas.

Pulsisure, mis on liiga madal (alla 30 mmHg) võib näidata südame pumpamisvõime nõrkust või muid meditsiinilisi probleeme.

Allikas: <https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings>

- *BS BodySugar* - veresuhkur.

Raseduse ajal veresuhkru norm 5,2–6,9 mmol. Kõrgem veresuhkur võib tähendada rasedusdiabeedi riski.

Allikas: <https://www.kliinikum.ee/patsiendiinfo-andmebaas/rasedusdiabeet-ja-glukoosi-taluvuse-test-gtt/>

- *BodyTemp* – kehatemperatuur.

Raseduse ajal normaalne kehatemperatuur umbes 99°F - 100°F.

- *HeartRate* – Pulss. Raseduse ajal tõuseb pulsi sagedus 10-15 löögi võrra minutis võrreldes raseduse eelse tasemega.

Raseduse ajal jääb normaalne pulss vahemikku 70-90 lööki minutis.

- *RiskLevel* – riski hinnang tervisele – madal, keskmine, kõrge.

1.8 Kulud ja kasud

Kuluks loeme projekti meeskonna ajakulu (hinnanguliselt 66 tundi) ja kasuks õppimine, areng ning uued põnevad teadmised.

1.9 Data-mining eesmärgid

See etapp jääb ära, kuna andmed on juba kogutud.

1.10 Data-mining edu kriteerium

See etapp jääb ära, kuna andmed on juba kogutud.

2. Data Understanding

2.1 Andmete kogumine

Andmed on juba eelnevalt kogutud. Kasutame kõiki andmestikus antud andmeid, tunnuseid ära ei jäta ja valimit koguandmetest ei koosta. Siiski, kui esineb ilmselgeid andmevigu, jätame endale õiguse neid korrigeerida.

2.2 Andmete kirjeldamine

Andmestik koosneb 1014 reast ja 7 veerust, millest 2 on ujukomaarvude tüüpi, 4 täisarvude tüüpi ja 1 objekti tüüpi. Puuduvaid väärtusi ei ole.

Column	Dtype	Attribute	Ühik
Age	int64	Continuous	Aastad
SystolicBP	int64	Continuous	mmHg (millimeetrit elavhõbedasammast)
DiastolicBP	int64	Continuous	
BS	float64	Continuous	mmol
BodyTemp	float64	Continuous	Fahrenheit
HeartRate	int64	Continuous	Löögid minutid
RiskLevel	object	Categorical (Low, Med, High)	Kategooria (Low, Med, High)

2.3 Andmete uurimine

Kirjeldame andmestiku tunnuste lihtsaid statistilisi näitajaid.

1. Vanus

Kõige noorem ema on 10-aastane, eakaim 70-aastane, keskmine vanus ligi 30 aastat. Standardhälve 13.47 näitab üsna suurt varieeruvust emade vanuses.

2. Ülemine vererõhk

Kõige kõrgem ülemise vererõhu näitaja on 160, mis viitab haiguslikule näitajale, 2. astme hüpertensioonile. Kõige madalam ülemise vererõhu näitaja on 70, mis viitab samuti ebanormaalselt madalale näitajale. Keskmine on 113.20, mis viitab täiesti tervele näitajale (optimaalne <120). Standardhälve on 18.40.

3. Alumine vererõhk

Kõige kõrgem alumise vererõhu näitaja on 100, mida loetakse kõrgeks (>90), kõige madalam alumise vererõhu näitaja on 49, mis tundub esmapilgul liiga madal näitaja. Keskmine alumise vererõhu näitaja on 76, mis viitab normaalsele tulemusele. Standardhälve on 13.89.

4. Veresuhkur

Kõige kõrgem veresuhkru näitaja on 19, madalaim 6, keskmine 8.73 ja standardhälve 3.30. Kõige kõrgem veresuhkru näitaja on väga kõrge, millesse oleks vaja edasises analüüsis sisse vaadata. Kõige madalm näitaja jääb rasedusaegse veresuhkru normi piiridesse, keskmine, aga ületab selle (kuni 6,9mmol) ja viitab rasedusdiabeedi kõrgemale riskile. Standardhälve on antud näitaja kohta võrdlemisi suur, seega esineb üsna suurt varieeruvust rasedate naiste veresuhkru näitajas.

5. Kehatemperatuur

Kõige kõrgem kehatemperatuuri näitaja on 103, kõige madalam 98, keskmine 98.67 ning standardhälve 1.37. Raseduseaegne normaalne kehatemperatuur jääb vahemikku 99°F - 100°F. Seega, antud andmestiku naised on keskmiselt normist veidi madalama kehatemperatuuriga. Esineb ka erindeid, nt naine kehatemperatuuriga 103 näol, mis viitab ilmselgelt palavikule.

6. Pulss

Kõige kõrgema pulsiga näitaja on 90, kõige madalam 7, keskmine 74.30 ja standardhälve 8.09. Maksimaalne ja keskmine pulss jäävad normi piiresse, kuid kindlasti peame antud näitajasse andmete puhastamise faasis sisse vaatama vaadates minimaalset näitajat, mis tundub olevat viga.

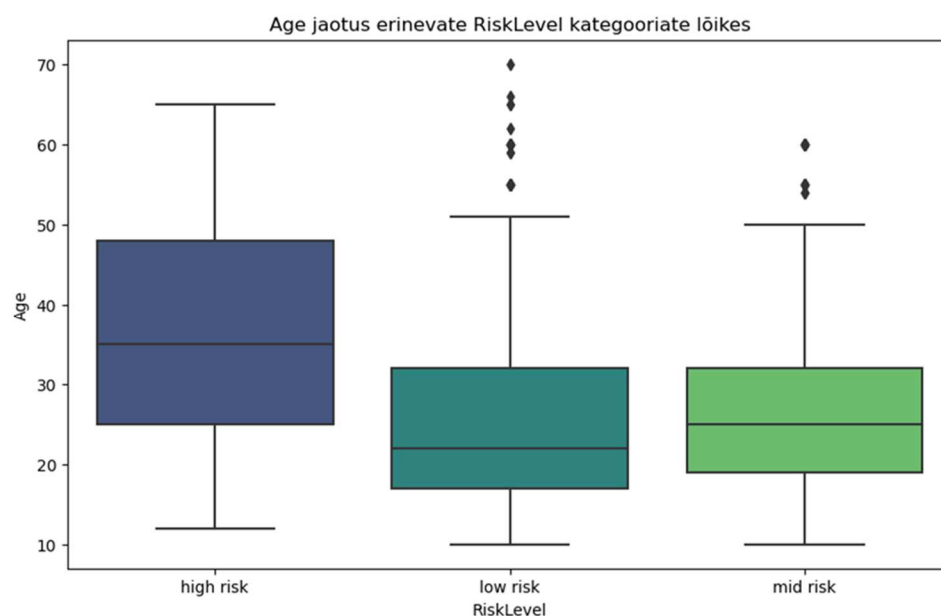
	Vanus	Ülemine vererõhu näitaja	Alumine vererõhu näitaja	Veresuhkur	Kehatemperatuur	Pulss
	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate
count	1014,0000	1014,0000	1014,0000	1014,0000	1014,0000	1014,0000
mean	29,8717	113,1982	76,4606	8.725986	98,6651	74,3018
std	13,4744	18,4039	13,8858	3,2935	1,3714	8,0887
min	10,0000	70,0000	49,0000	6,0000	98,0000	7,0000
max	70,0000	160,0000	100,0000	19,0000	103,0000	90,0000

Selgitame esialgu välja erinevate tervisenäitajate mõju riskihinnangule lihtsustatud meetodil, et saaksime esialgse pildi ette. Projekti edasistes osades süveneme täpsemalt erinevate tunnuste mõjudele ja koosmõjudele nt madala ja keskmise riski erinevustele.

2.4 Tervisenäitajate mõju riskihinnangule

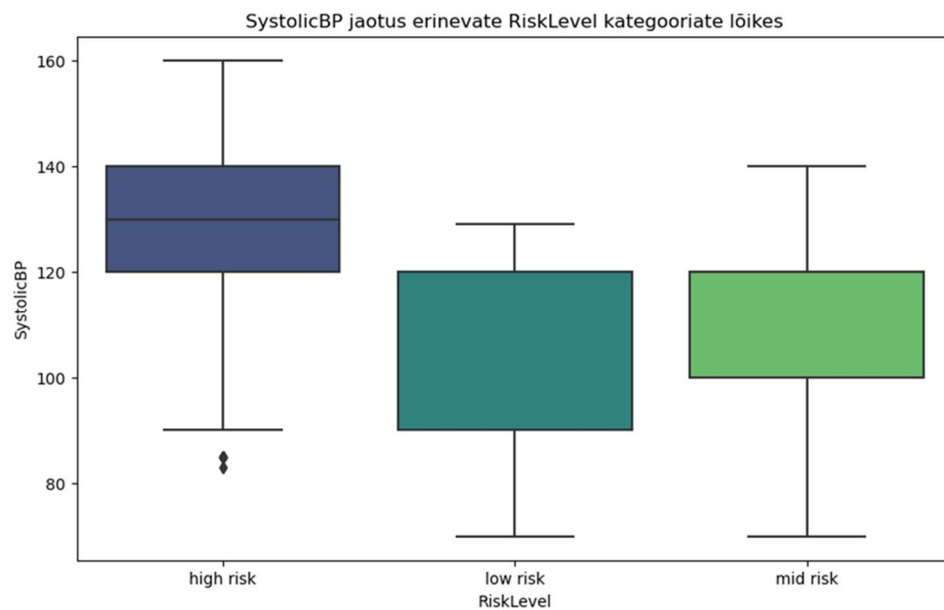
1. Vanus

Näeme allolevatelt karpdiagrammidelt, et kõrgem risk tundub olevat mõjutatud vanusest, st. mida kõrgem vanus seda kõrgem risk. Madala ja keskmise riskigrupiga naiste vanused tunduvad võrdlemisi sarnased. Nii madalas kui keskmise riskigrupiga naiste vanustest on ka erindeid. Kindlasti tasub täpsemalt analüüsida ligi 70 aastaseid naisi, kelle vanus antud teema kontekstis tundub ebanormaalne. Küll aga same järeldada, et kõrgemas riskiklassis olevate naiste vanus varieerub rohkem kui madalas ja keskmises riskiklassis.



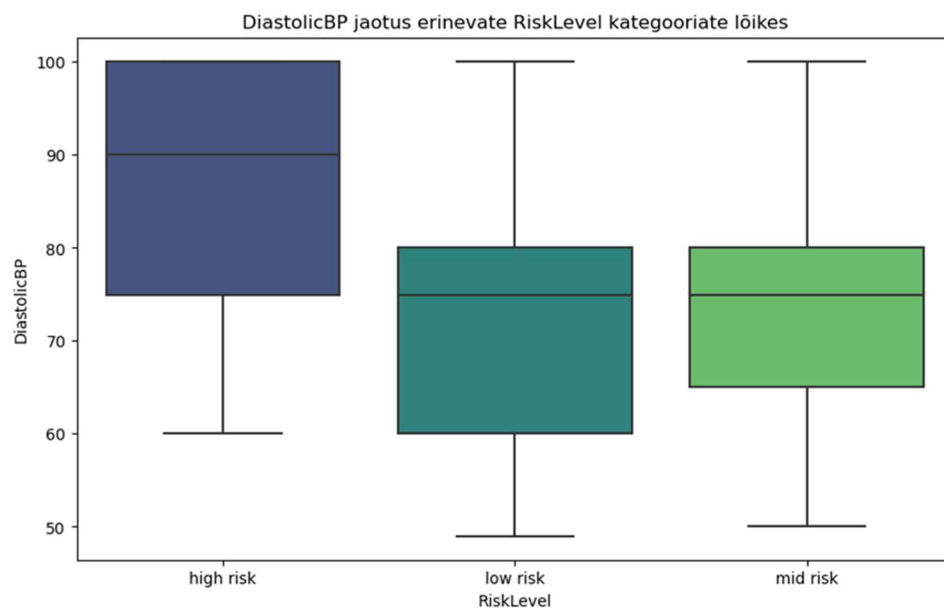
2. Ülemine vererõhk

Näeme allolevatelt karpdiagrammidelt, et ülemine vererõhk mõjutab riskihinnangut. Eriti hästi on seda näha kõrge riskiga naiste puhul, et nende ülemine vererõhk on tunduvalt kõrgem keskmise ja madala riskihinnanguga naistest. Keskmise ja madala riskihinnanguga naiste tulemused tunduvad esmapilgul küllalt sarnased. Ka siin esinevad mõned erindid ning madala ja keskmise riskihinnanguga naiste puhul ei ole toodud ka ülemise vererõhu mediaani.



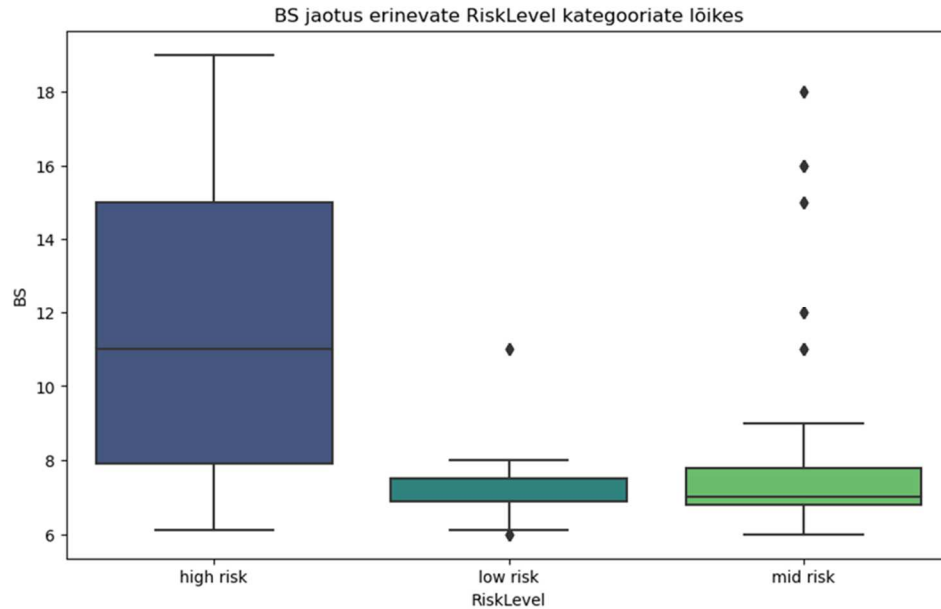
3. Alumine vererõhk

Näeme allolevatelt karpdiagrammidelt, et riskihinnangut võiks mõjutada ka alumine vererõhk, eriti kõrgemat riskihinnangut, kus naistel on ka kõrgem alumine vererõhk. Keskmise ja madala riskihinnanguga naiste alumised vererõhud on taaskord võrdlemisi sarnased.



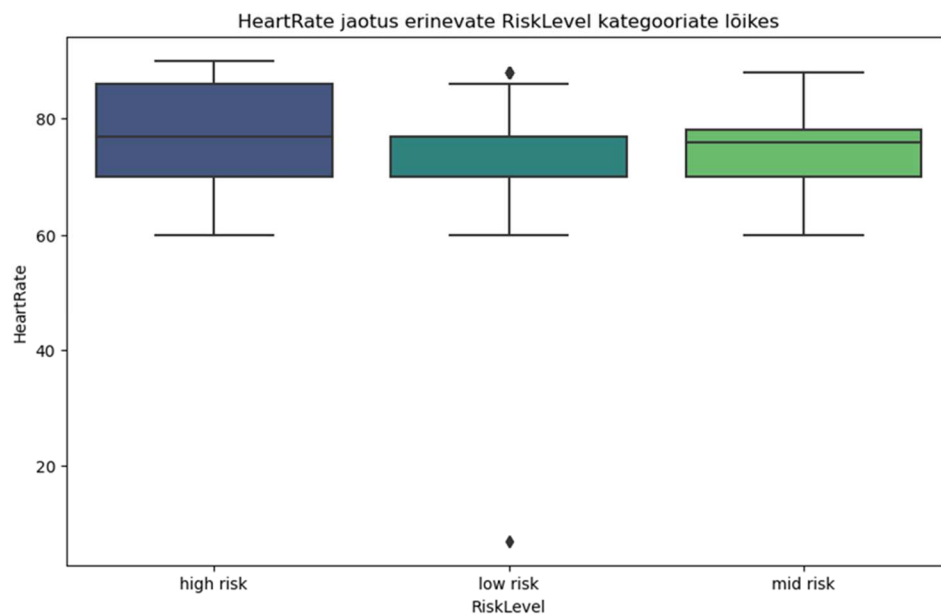
4. Veresuhkur

Näeme allolevatelt karpdiagrammidelt, et veresuhkur mõjutab riskihinnangut, kus kõrgema veresuhkruga naistel on ka kõrge riskihinnang. Keskmise ja madala riskihinnanguga naiste veresuhkru tulemused on taaskord võrdlemisi sarnased, aga esineb erindeid.



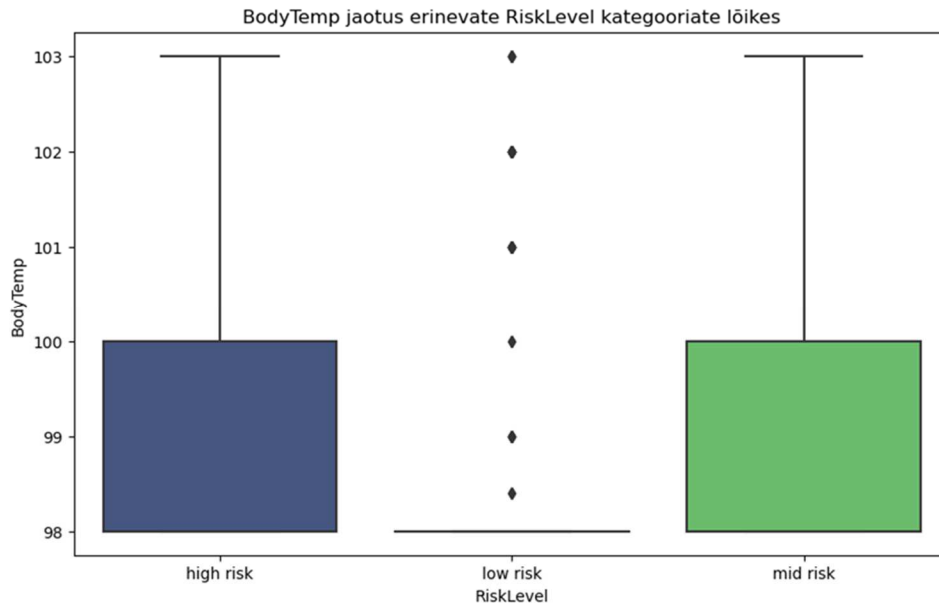
5. Pulss

Näeme allolevatelt karpdiagrammidelt, et pulsil ei tundu esmapilgul olevat nii suurt mõju riskihinnangule. Diagrammidelt on küll näha, et esineb üks väga madala pulsiga erind, millega oleks kindlasti puhastamise faasis mõistlik tegeleda.



6. Kehatemperatuur

Näeme allolevatelt karpdiagrammidelt, et kehatemperatuuril ei paista olevat olulist mõju riskihinnangule, kusjuures kõrge ja keskmise riskihinnanguga kehatemperatuuride jaotused on väga sarnased ning madala riskihinnanguga kehatemperatuurid oleks justkui kõik erindid. Madala riskihinnanguga kehatemperatuuridesse võiks analüüsi käigus veel sisse vaadata.



2.5 Andmete kvaliteet

Esinevad väiksed sisestamise-mõõtmise vead, kus mõnel juhul on nt võimatult madal pulss jne. Sellised kvaliteedivead on parandatavad ja tegeleme nendega andmete puhastamise käigus.

3. Projekti planeerimine

Projekti plaan:

1. Teema valik ja eelanalüüs
 - a. Hannes 4h
 - b. Natalja 4h
 - c. Mari 4h
2. Andmete puhastamine
 - a. Hannes 1h
 - b. Natalja 1h
 - c. Mari 1h
3. Andmetele lisandväärtuse tekitamine (vajadusel uued tunnused, andmete liitmine ja format)

- a. Hannes 2h
 - b. Natalja 2h
 - c. Mari 2h
4. Mudelite valimine ja testimine – katsetame millised mudelid meie püstitatud eesmärki aitaksid täita
- a. Hannes 6h
 - b. Natalja 6h
 - c. Mari 6h
5. Valitud mudelite loomine ja hindamine
- a. Hannes 2h
 - b. Natalja 2h
 - c. Mari 2h
6. Tulemuste hindamine
- a. Hannes 2h
 - b. Natalja 2h
 - c. Mari 2h
7. Tehtud protsessi ülevaatamine (kas jäi midagi olulist kahe silma vahele)
- a. Hannes 1h
 - b. Natalja 1h
 - c. Mari 1h
8. Lõplike tulemuste raporteerimine
- a. Hannes 2h
 - b. Natalja 2h
 - c. Mari 2h
9. Lõplike tulemuste üle vaatamine
- a. Hannes 1h
 - b. Natalja 1h
 - c. Mari 1h
10. Ettearvamatud tööd
- a. Hannes 1h
 - b. Natalja 1h
 - c. Mari 1h