

Tartu Ülikool
Sotsiaalteaduste valdkond
Psühholoogia Instituut

Karl Martin Kadaja
Söömishäirete küsimustike väiksemaks tegemine masinõppega ristvalideerimise kaudu
Uurimistöö

Juhendajad: Uku Vainik, Kadri Arumäe
Läbiv pealkiri: Küsimustike lühendamine

Tartu 2025

Kokkuvõte

Söömishäirete küsimustike väiksemaks tegemine masinõppega ristvalideerimise kaudu

Antud uuringu eesmärk oli leida võimalikult lühike, kuid kehamassiindeksit (KMI) täpselt ennustav komplekt, mis tuleb kokku 318 küsimusega 11-st erinevast emotsiooni- ja söömishäirega küsimustikust. 990 Põhja-Ameerika õpilast (539 naist, 451 meest) jagati 80% treeningandmeteks ja 20% testandmeteks. Seejärel jooksutati nii Lasso mudelit, kui ka RFECV meetodit koos järgnevate mudelitega: mitmene regressioon, otsustuspuu ja XGBoost. Samuti häälestati ka hüperparameetrid korduva *KFold* ristvalideerimisega. XGBoost andis kõige paremaid tulemusi testandmestikuga valideerimisel (naistel $R^2=0.280$, meestel $R^2=0.273$). Stabiilsuslävendiga $p > 90\%$ valiti XGBoosti mudeli leitud tunnustest 57 sobivat meeste küsimust ja 37 naiste küsimust. Meeste KMI'd ennustasid peamiselt DEBQ ja AUDIT küsimustikud, naistel oli variatiivsus suurem. Masinõpe on võimeline vähendama tunnuste arvu selliselt, et KMI ennustusvõime säilib tugev.

Märksõnad: masinõpe, ristvalideerimine, söömishäired, kontrollimata söömine, KMI

Abstract

Reducing Eating Disorder Questionnaires via Machine Learning and Cross-Validation

The purpose of the study is to find a short yet accurate item set for predicting body-mass index (BMI) from 318 items across 11 eating- and emotion-related scales. Data from 990 North-American students (539 women, 451 men) were split 80/20. Lasso and recursive feature elimination with cross-validated linear regression, random forest and tuned XGBoost were compared. Using a $\geq 90\%$ stability threshold, XGBoost retained 57 male and 37 female items and yielded the highest hold-out R^2 (women 0.280; men 0.273). Male prediction relied mainly on DEBQ and AUDIT items, whereas female signal spanned six scales. Machine-learning feature selection can roughly halve questionnaire length without losing much predictive accuracy.

Keywords: machine learning, cross-validation, eating disorders, uncontrolled eating, BMI

Sissejuhatus

Hiljuti on hakatud tegema üha rohkem söömishäirete kohta küsimustikke. Pikka aega on selles valdkonnas loodud väga erinevate nimedega küsimustikke - *Binge eating scale*, *Loss of control over eating scale* jne. Küsimustike psühhomeetrilised omadused on väga head (nt viimasel sisemine koosõla $\alpha = .88$, Racine, 2019), kuid need on väga mahukad ning neil on üle erinevate küsimustike sarnane sisu (Vainik U. et al., 2019).

Sageli tekib pikkade küsimustike vastajatel väsimus, nende tähelepanu hajub. Seetõttu nad kas kiirustavad või ei mõtle küsimustele piisavalt ning vastused tulevad ebatäpsed. Vähendamise abil säästaksime inimeste vaeva ja võimalik, et parandaksime nende motivatsiooni küsimustikke täita. Samuti analüüsime sellega ka erinevate valdkondade kattuvust. Küsimustik võiks olla üsna paindlik, et ennustada söömishäirete aspekte, mis võiks hõlmata ka vastavaid sümptomeid, kehamassiindeksit (edaspidi KMI) ning emotsioonide reguleerimise probleeme.

Selle töö aluseks võtame autori Sarah E. Racine 2019. aasta uuringu, milles olevad andmed on saadud erinevate küsimustike vastuste põhjal. Meie töö annab siinkohal ülevaate mudeli loomise uutest võimalustest. Samuti ei tohi küsimuste lühendamisega teha kahte tüüpilist viga. Esiteks ei tohi uskuda, et küsimustiku lühendamisega selle usaldusväärsus ja valiidsus kattub täpselt täispika küsimustikuga. Teiseks ei tohi uskuda, et kuna küsimusi on vähem, siis on ka valiidsed andmeid vähem (Smith G. T. et al., 2000). Juba varasemalt tehtud uuringud näitavad, et juba poole vähemate küsimustega küsimustikud korreleeruvad tugevasti algsete küsimustikega (Vainik U. et al., 2019). Järelikult võiks luua ühe suure konstrukti: kontrollimata söömine (Vainik U. et al., 2019). Selle konstruktiga ennustame KMI-d.

Me püüame luua vähendatud nn võimalikult hästi KMI-d ennustava küsimustiku, kasutades erinevalt Racine et al aritklist masinõpet. Psühholoogide jaoks on tavaline kasutada regressiooni - ehk teha andmestiku pealt üks mudel. Meie üritame luua vähendatud andmestiku osa pealt mudeleid, mis ennustaksid välja jäänud andmestikku. See annab vihjeid, kui hästi võiks mudel ka teiste andmestikega töötada (Yarkoni et al., 2017).

Kokkuvõttes oleks efektiivne luua küsimustik, mis oleks esialgsele küsimustikule võimalikult sarnase mudeliga, kuid piisavalt väheste küsimustega. Selleks analüüsime masinõppe kaudu nelja mudeli abil küsimuste vähendamise võimalusi. Viimasel ajal on teadustöodes üha rohkem kasutama hakatud rekursiivset tunnuste vähendamise meetodit (Priyatno et al., 2024).

Üritame töös kasutada antud meetodit koos mitmese regressiooni, otsustuspuu ja ekstreemse gradientvõimendamise mudelitega, et vähendada ebaolulisi küsimusi ja müra. Lisaks testime ka Lasso võimekust tunnuste vähendamisel. Uurimistöö eesmärk on leida antud mudelite statistiline olulisus vähendatud tunnustega KMI ennustamisel. Samuti oleme huvitatud, milline erinevus on lineaarsetel ja otsustuspuudel põhinevate mudelite tulemuste vahel. Viimaks soovime teada, milline on leitud vähendatud tunnuste kirjeldus ja erinevus nii meeste kui ka naiste andmestikus.

Meetod

Ülevaade lähteandmetest ja eeltötlusest

Valim

Valim on saadud Racine 2019. aasta uuringust, kus 998 inimest olid vastanud söömishäirete küsimustikkudele. Peale eeltöötlemist kasutati mudelite treenimisel 539. naise ja 451. mehe andmeid, kelle vanused jäid vahemikku 18-44 aastat(Tabel 1).

Tabel 1

Uuringus osalejate arv vanusegruppide kaupa

Vanusegrupp	Naised		Mehed	
	n	%	n	%
18-20	476	48.08	366	36.97
21-25	60	6.06	75	7.58
26-30	1	0.1	7	0.71
31-35	1	0.1	1	0.1
36-40	1	0.1	1	0.1
41-45	0	0	1	0.1
Kokku	539	54.44	451	45.66

Küsimustikke täitsid vähemalt 18-aastased Põhja-Ameerika haridusinstituutsiooni psühholoogiatudengid. Andmete valiidsete väärtuste saamiseks oli küsimustikus 3 kontrollküsimust. Algandmestikust eemaldati osalejad, kes täitsid küsimustiku ebarealistliku kiirusega või ebatavalise kordava vastamismustriga.

Küsimustike tunnused

Ennustavaid tunnuseid oli algselt kokku 351, mis tulid 11-st erinevast söömishäirete ja emotsioonide reguleerimisega seotud küsimustikust. Nende vahemikud olid täisarvud skaaladel 1-2 kuni 1-8. Küsimustikud Racine 2019 aasta tööst olid järgmised:

1. DEBQ – Hollandi söömiskäitumise küsimustik (ingl k. *Dutch Eating Behavior Questionnaire*) sisaldab 13 küsimust, mis mõeldavad inimeste söömiskäitumist pärast negatiivsete emotsioonide kogemist (van Strien et al., 1986)
2. EPSI – Söömishäirete sümptomite loend (ingl k. *Eating Pathology Symptoms Inventory*) sisaldab 45 küsimusega liig-söömisega seotud alaskaalat (Forbush et al., 2013)
3. LOCES – Söömise üle kontrolli kaotamise skaala (ingl k. *Loss of Control over Eating Scale*) sisaldab 24 küsimust kontrollimatu söömise üle, mis ei sisalda koguseid arvuliste väärtuste näol (Latner et al., 2014)
4. BES – Liigsöömise skaala (ingl k. *Binge Eating Scale*) sisaldab 16 küsimust, mis uurivad emotsionaalseid osasid liig-söömisest (Gormally et al., 1982)
5. EDEQ – Söömishäirete hindamise küsimustik (ingl k. *Eating Disorder Examination Questionnaire*) on küsimustik, kui inimene ise hindab enda söömishäirete sümptomeid (Fairburn et al., 1994)
6. PFS – Toidu mõju skaala (ingl k. *Power of Food Scale*) on 15 küsimusega ise hinnatav küsimustik, kus uuritakse meeldiva toidu mõju söömiskäitumisele (Lowe et al., 2009)
7. IDAS II – Depressiooni ja ärevuse sümptomite küsimustik II (ingl k. *Inventory of Depression and Anxiety Symptoms II*) on 99 küsimusega küsimustik, millega saab mõõta tujusid, ärevust ja mõnu mitte tundmist (Watson et al., 2012)
8. AUDIT – Alkoholi tarvitamise häirete tuvastamise test (ingl k. *Alcohol Use Disorders Identification Test*) sisaldab 10 alkoholi joomisega seotud küsimust (Saunders et al., 1993)
9. DERS – Emotsioonide regulatsiooni raskuste skaala (ingl k. *Difficulties in Emotion Regulation Scale*) mõõdab 36 küsimuse abil emotsionaalse selguse, uimasuse, vastuvõtlikkuse ja implusiivsuse näidikuid (Gratz et al., 2004)
10. CIA – Kliinilise kahjustuse hindamise küsimustik (ingl k. *Clinical Impairment Assessment*) sisaldab 16 küsimust, mis mõeldavad viimase 28 päeva söömishäiretest tingitud sümptomeid (Bohn et al., 2008)

11. YFAS – Yale'i toidusõltuvuse skaala (ingl k. *Yale Food Addiction Scale*) mõõdab sõltuvust toidust viimase aasta jooksul. See 25 küsimusega skaala põhineb DSM-IV-TR sõltuvuse kriteeriumidel (Gearhardt et al., 2009)

Andmeanalüüs

Eeltöötlemine

Esialgne andmestik sisaldas 998 rida ning 362 veergu. Puhastamise käigus eemaldati 8 rida ning 44 veergu. Kuna uuring tehti noorte ülikooliõpilaste seas, siis on 93.3(3) protsenti vastanutest 18-21 aastased (Tabel 1). Antud juhul jääb vanus KMI ennustamisel statistiliselt oluliseks ($p=0.11$) ning sõltumatu muutujana seda ei eemaldatud. YFAS küsimustikus oli puudu 217 vastust. Eemaldasin selle küsimustiku algandmestikust, sest kogu valim oli väike ($n=990$) ning see oleks tekitanud ebaolulist müra.

Samuti on algandmestikust välja jäetud seitse EDEQ'i küsimust (edeq17, edeq18, edeq20, edeq22, edeq24, edeq26, edeq28) ja üks AUDIT'i küsimus (audit11), sest need olid kas eelmisele küsimusele järgnevad või fikseerimata vahemikuga küsimused. Seejärel teisendati kõik väärtused ujukomaarv-tüübile ning eemaldati kõik *NaN* väärtused. Pärast viimaseid meetodeid sisaldas andmestik 990 rida.

Algandmestiku teatud vastajatel olid EPSI küsimustiku väärtused ühe võrra nihkes. Selle ebatäpsuse parandamiseks suurendati vastavate ridade väärtuseid. Peale andmete puhastamist jäi lisaks KMI-le, soole ja vanusele alles 318 küsimustiku tunnust. Andmestiku kõik väärtused arvutati ümber nii, et kõik vahemikud viidi kujule 0-1 (välja arvatud KMI) ning salvestati hilisemateks mudeldamiseks eraldi faili.

Statistiline analüüs

Instrumentideks on kehamassiindeks, söömishäired ja emotsioonide reguleerimise probleemid. Uuringu sõltuv muutuja on kehamassiindeks (KMI), mis arvutatakse valemi abil: $KMI = \text{kaal}/(\text{pikkus}*\text{pikkus})$. Uuringu sõltumatud muutujad on sugu, vanus ja küsimustike vastused. Sugu on binaarne muutuja: mees või naine. Vanus on arvuline tunnus ning mõõdetud aastates.

Jagame andmed kaheks. 80% andmestikust läheb mudeli treenimiseks ja 20% selle väliseks valideerimiseks. Mudeleid luuakse hüpoteesivabalt, üritame rekursiivse tunnuste elimineerimise võttega pimesi ennustada parimat vähendatud tunnuste valimit. Kasutame

tulemuste leidmiseks eraldi Lassot ja siis RFECV-d koos järgmiste mudelitega: mitmene regressioon, peenhäälestatud otsustusmets ja peenhäälestatud ekstreemne gradientvõimendamine (tutvustatud allpool). Nagu on välja toodud soo lõikes erinevused Racine 2019 aasta töös, siis naiste ja meeste andmeid treenitakse eraldi, et säilitada täielik faktorstruktuur. Peale optimaalsete tunnuste leidmist soovime luua parima ligikaudse mudeli laialdaseks kasutuseks nägemata andmestiku peal. Valideerime mudeli välise test-andmestikuga (20%). See on vajalik selleks, et ei tekiks märkamatu andmeleket ning annab meile ülevaate mudeli täpsusest. Hindame mudeli R^2 ja RMSE väärtuseid ning teeme mudelitevahelisi analüüse. Selleks mõõdame mudelite ennustusvõimet ning tunnuste esinemist nii mudelite kui ka soo lõikes.

Rekursiivne tunnuste elimineerimine

Rekursiivne tunnuste elimineerimine (ingl k *recursive feature elimination*, edaspidi RFE) aitab järk-järgult eemaldada vähem olulisi küsimusi, jättes alles vaid kõige olulisemad tunnused. RFE jaoks valitakse parameetriteks mudel, tunnuste arv ja sammude suurus. Mudel võib olla mitmene regressioon, otsustuspuu või muu sobiv mudel. Tunnuste arv on vabalt valitud konstant 1-st kuni andmestiku maksimaalsete tunnuste arvuni. Sammude arv defineerib, mitu tunnust kaotatakse igal sammul. Iga samm kaotab kõige nõrgema ennustusvõimega tunnused. Mida rohkem kaotatakse tunnuseid, seda suurem võimalus, et kaob ka tegelikult oluline tunnus. Mida vähem kaotatakse tunnuseid, seda kauem võtab aga mudeli treenimine aega. (Guyon et al, 2002)

Rekursiivne tunnuste elimineerimine koos ristvalideerimisega

Rekursiivne tunnuste elimineerimine koos ristvalideerimisega (ingl k *recursive feature elimination with cross validation*, edaspidi RFECV) toimib sarnaselt RFE-le rekursiivse tunnuste elimineerimisega, kuid lisaks on juures ka iga-sammuline ristvalideerimine. Parameetriteks valitakse mudel, sammu suurus, ristvalideerimiste arv, skoori meetod ja minimaalsete tunnuste arv. Skoori meetod võib olla r^2 , *negative_root_mean_squared_error*. Iga RFECV samm tehakse ristvalideerimine ja saadakse defineeritud skoor. Lõpuks valitakse mudel tunnuste arvuga, mille ristvalideerimisest saadud skoor oli kõige kõrgem.

Lasso

Lasso (ingl k. *Least Absolute Shrinkage and Selection Operator*) on regressioonimudel, mis pidevalt valib tunnuseid, samal ajal nõrgemaid lüüsid ja nende koefitsente nullini kahandades. See tähendab, et vähem tähtsad tunnused elimineeritakse ning ülejäänud tunnuste struktuur ennustab hästi ennustatavat väärtust. (Tibshirani et al., 1996)

Otsustusmets

Otsustusmets (ingl k. *random forest*) on tugev masinõppe mudel, mis valib juhuslikke osi andmestikust, et treenida üksikuid otsustuspuuid. Erinevalt üksikust otsustuspuust, kus otsuseid tehakse ühe puu struktuuri põhjal, võtab otsustusmets kokku mitme puu ennustused, et teha lõplik otsus. Otsustusmets on väga hea üldistusvõimega ning suudab mudeldada keerulisemaid seoseid sisendite ja väljundite vahel. Lisaks suudab otsustusmets käsitleda puuduvaid väärtuseid, mis võtab ära vajaduse sisestada puuduvaid andmeid. (Sügis et. al., 2024; Breiman et al., 2001)

Ekstreemne gradientvõimendamine

Ekstreemne gradientvõimendamine (ingl k. *extreme gradient boosting*, edaspidi XGBoost) on otsustuspuudel põhinev keeruline algoritm, millele on sisseehitatud meetodid otsustamiseks, mida teha puuduvate väärtustega, ja suutlikkus tõhusalt kasutada olemasolevat riistvara (paralleelarvutused, vahemälu kasutamine), et mudel kiiremini õpetatud saaks. Samuti on arvesse võetud ülesobitamise vastaseid meetodeid ja otsustuspuude lõikamise meetodeid ning iga mudeli loomiseks kasutatakse ristvalideerimist. (Sügis et. al., 2024; Chen et al., 2016)

Mudeli peenhäälestamine

Mudeli loomisel kasutatakse tihti hüperparameetreid. Need on väärtused, mis fikseeritakse enne mudelite treenimist ning millest sõltub mudeli ennustusvõime. Teatud olukordades võib ennustusvõime hüperparameetrite väärtustest oluliselt sõltuda. Hüperparameetrite katsetamist nimetatakse peenhäälestamiseks ning seda tehakse nii kaua, kuni tulemus on kõige sobivam (Sügis et. al., 2024).

Ennustamise meetod

Mudelite ennustusvõime hindamiseks ning tulemuste reprodutseerimiseks järgitakse mudelite treenimisel, et metoodiline protsess püsiks muutumatu. Tänu sellele säilib tulemuste valiidsus ka uute mudelite jooksumisel, mis annab meile võimalikult täpse tulemuste analüüsi. Selle jaoks luuakse eraldi seemnete vahemikud. Seeme annab meile võimaluse juhuslikku protsessi reprodutseerida. See võib olla näiteks andmestiku poolitamine treening- ja testandmeteks või otsustusmetsa puu arvutamise samm. Testandmestiku vahemikuks valisime juhuslikult 0-9. Meie töös kasutatakse seemneid kõikide poolitamiste ja mudeli jooksumise puhul.

Esmalt eemaldatakse KMI veerg, mis säilitatakse eraldi objektina, et hiljem anda edasi mudelile. Seejärel kasutatakse *for*-tsükli ning võetakse selle vahemikuks 0-9. Tsükli alguses jagatakse andmed kaheks, millest 80% läheb treeninguks ning 20% läheb väliseks testandmeks. Tsükli indeksit (0, 1, 2, ...) kasutatakse andmete poolitamiseks. Iga kord ühe mudelitüübi jooksumisel segatakse seemne abil andmestik ja valitakse 539 naise või 451 mehe seast väliseks test-valimiks vastavalt 108 naist või 90 meest. Treenimiseks jääb 431-st naisest ja 361-st mehest koosnevad andmestikud. Peale treening-test andmete saamist minnakse uude, sisemisse tsükliks, mille vahemikuks võeti juhuslikult 0-2. Need tsükli indeksid ehk seemned määravad RFECV meetodi juhusliku oleku väärtused, mudelite treenimisväärtused ning ristvalideerimiste väärtused. Mõlema, nii välise, kui ka sisemise tsükli indeksite kasutamine tagab tulemuste kordamise võimaluse.

Sisemises tsükliks kasutatakse kõigepealt tunnuste vähendamise meetodit RFECV-d. Selle jaoks valitakse esmalt välja meetodisene ennustusmudel (ingl k. *estimator*). Seejärel valitakse lisaparameetritena ristvalideerimise meetod. Valisime selleks fikseeritud *RepeatedKfold* validaatori poolitusarvuga 5 ja korduste arvuga 3. Selle abil jagatakse andmestik viieks ning tehakse RFECV igal sammul ristvalideerimisi. Tasub märkida, et *RepeatedKfold*'ile määrati juhusliku segamise seeme, mis on sama mudeli juhusliku seemnega. Skoorimise parameeter valiti RFECV jaoks negatiivne keskmine ruuthälve, sest see on täpsem ja vähem volatiilsem kui r -ruuduga skoorimine. Veel valitakse RFECV parameetriteks: minimaalsete tunnuste arv, millest alla tunnuseid ei ennustata; samm, üle mitme tunnuse korratakse ristvalideerimist; juhusliku oleku väärtus. Lasso puhul RFECV-d ei kasutata, sest selle loogika on tunnuseid vähendav. Lasso puhul kasutatakse sama *RepeatedKfold*'i, mida kasutati eelmiste mudelite puhul.

Ühel treenimise analüüsil kasutatakse ennustusudelina järgnevast kolmest erinevast mudelist ühte - mitmene regressioon, otsustusmets või XGBoost. Peale RFECV vastuste kätte saamist sobitatakse optimaalsed tunnused koos välise test-andmestikuga samale mudelitüübile, mis oli RFECV ennustusudel, kuid millele leiti peenhäälestamise kaudu veel lisaks parimad hüperparameetrid. Nii saadakse lõplik mudel. On väga oluline, et RFECV sisemine ennustusmeetod ja lõplik ennustusudel on sama tüüp, kas lineaarne või puu-baasil, kuna nende koos kasutamisel on tulemused ebatäpsed. Selle tõttu jooksutasime kõiki mudeleid eraldi peenhäälestamisega, et tulemus säiliks stabiilne.

Tulemused

Mudelite tugevus

Analüüsisime nelja erinevat mudelit - Lasso ning RFECV koos mitmese regressiooni, otsustuspuu ja XGBoostiga. Leidsime esmased (ristvalideerimise keskmised) ja hilisemad (testandmestikuga sobitamise) R^2 väärtused (vt. Tabel 2).

Leitud optimaalsete tunnuste arvu ja R^2 vaheline seos ei ole pidev. Puu-põhised mudelid saavutasid kõrge tulemuse ka suurte tunnuste arvuga (XGBoost $n=199$ esmane $R^2=0.353$, hilisem $R^2=0.232$) (vt. Joonis 2 ja Joonis 3). Tunnuste stabiilsus oli parim otsustusmetsal (keskmine välimine $R^2=0.257$, $STD=0.089$, $RMSE=3.849$). XGBoost oli sellele lähedal (keskmine välimine $R^2=0.276$, $STD=0.105$, $RMSE=3.794$). Lineaarsed mudelid said tulemusi madalate tunnuste arvu juures, kuid nende väline R^2 oli madalam (Lasso $R^2=0.177$, mitmese regressiooni $R^2=0.078$) (vt Joonis 4 ja Joonis 5).

Tulemused seemnete kaupa kõikusid -0.2 ja 0.42 vahel (vt. Joonis 1). Kõige stabiilsemalt kõrgemad R^2 väärtused saavutas nii meeste- kui ka naiste valimis XGBoost (naistel esmane $R^2=0.270$, hilisem $R^2=0.280$; meestel esmane $R^2=0.319$, hilisem $R^2=0.273$). Mitmene regressioon vastupidiselt oli nõrkade tulemustega (hilisem naistel $R^2 = 0.094$, meestel 0.062).

Tunnuste tulemused meestel ja naistel

Võrdlesime naise ja meeste vahelisi tunnuste korrelatsioone (Vt Tabel 2). Puu-põhiste mudelite põhjal on naiste keskmine algne tulemus $R^2=0.262$ ja hilisem tulemus $R^2=0.261$. Meestel oli vastavalt algne $R^2=0.312$ ja hilisem $R^2=0.273$.

XGBoosti hea tulemuse tõttu leiti stabiilsuslävendiga ($p > 90\%$) keskmine küsimustike tunnustepõhine andmestik, mis sisaldas 57 meeste ja 37 naiste stabiilset tunnust. Nendest kattuvaid oli 15 (vt Tabel 3). Naiste ja meeste ühised küsimused kattusid 20.7% (Jaccardi indeks). Meeste seas kerkisid esile küsimustikud DEBQ, EPSI, IDAS-II ja EDEQ, naiste seas aga EPSI, EDEQ, IDAS-II ja DERS. Detailset väljavõtet küsimustest on näha Lisa A ja Lisa B sektsioonis.

Tabel 2*Kõikide küsimuste filtreeritud meeste ja naiste ühisosa ja eraldiseisev osa*

Küsimustik	Ühisosa		Mehed		Naised	
	Tunnuseid	Protsent	Tunnuseid	Protsent	Tunnuseid	Protsent
DEBQ	2	13.333	12	28.571	1	4.545
EPSI	7	46.667	10	23.81	7	31.818
LOCES	0	0	0	0	1	4.545
BES	1	6.667	1	2.381	1	4.545
EDEQ	4	26.667	4	9.524	3	13.636
PFS	0	0	2	4.762	0	0
IDAS-II	0	0	7	16.667	4	18.182
AUDIT	0	0	3	7.143	1	4.545
DERS	0	0	2	4.762	3	13.636
CIA	0	0	1	2.381	1	4.545
YFAS	0	0	0	0	0	0
Vanus	1	6.667	0	0	0	0
Kokku	15	100	42	100	22	100

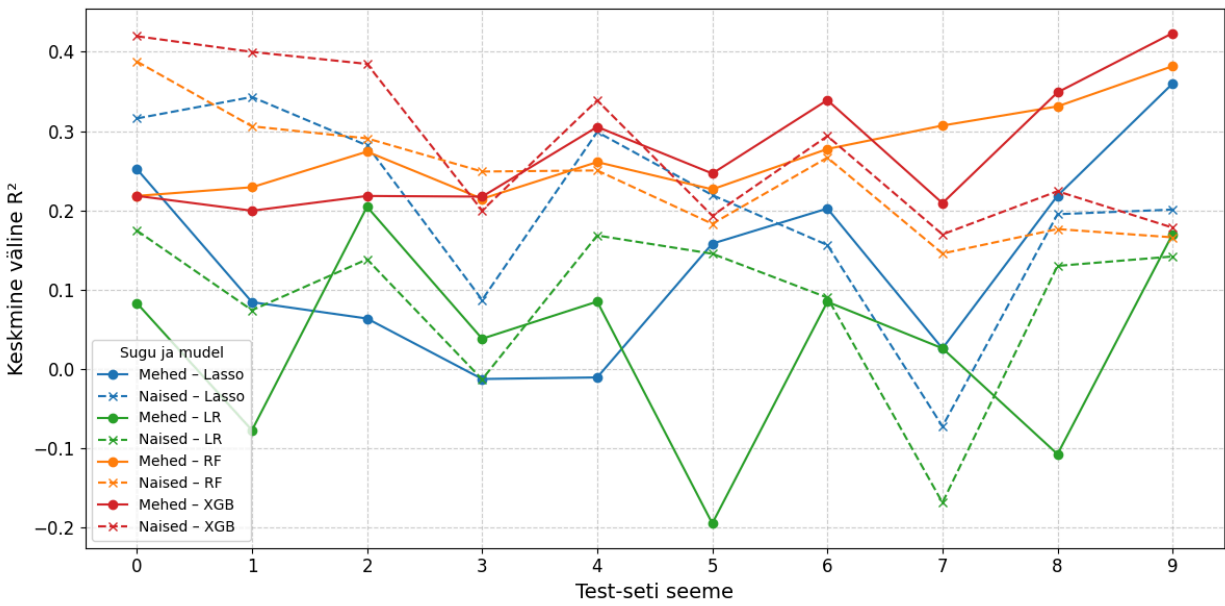
Tabel 3*Keskmsed algsed ja hilisemad R^2 tulemused.*

Mudel	Naised		Mehed	
	Algne R^2	Hilisem R^2	Algne R^2	Hilisem R^2
RFECV	0.250	0.088	0.313	0.031
Mitmene Regressioon				
Lasso	0.337	0.203	0.362	0.134
RFECV	0.254	0.242	0.305	0.272
Peenhäälestatud Otsustusmets				
RFECV	0.270	0.280	0.319	0.273
Peenhäälestatud XGBoost				
Kokku	0.278	0.203	0.325	0.178

Märkus. Visuaalse pildi jaoks vaata Lisa D.

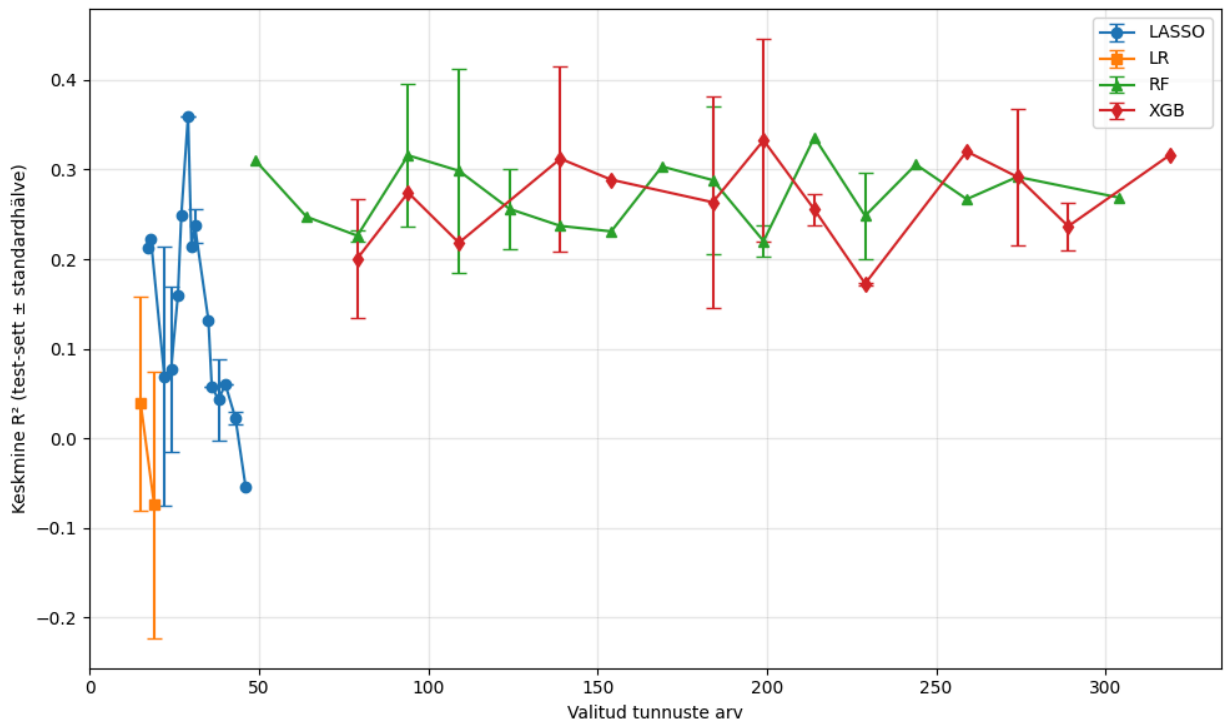
Joonis 1

Keskmine väline R^2 skoorid vastavalt testandmete seemnele



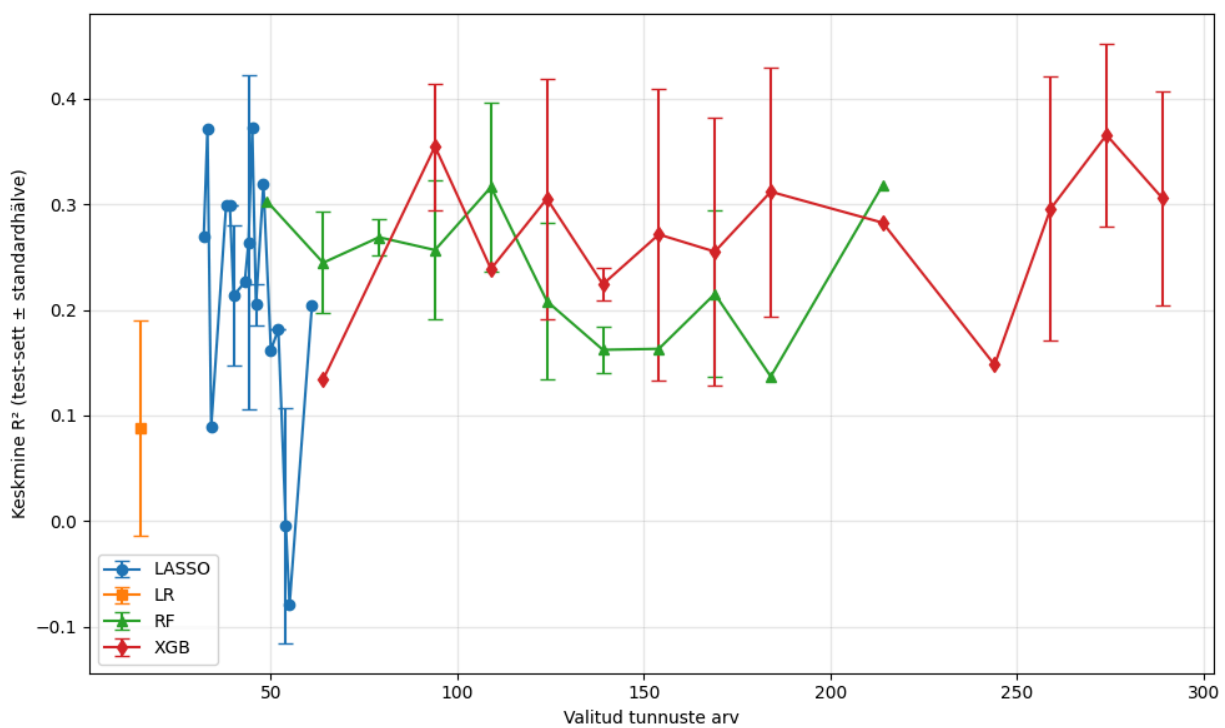
Joonis 2

Meeste keskmine väline R^2 väärtus vastavalt valitud tunnuste arvule



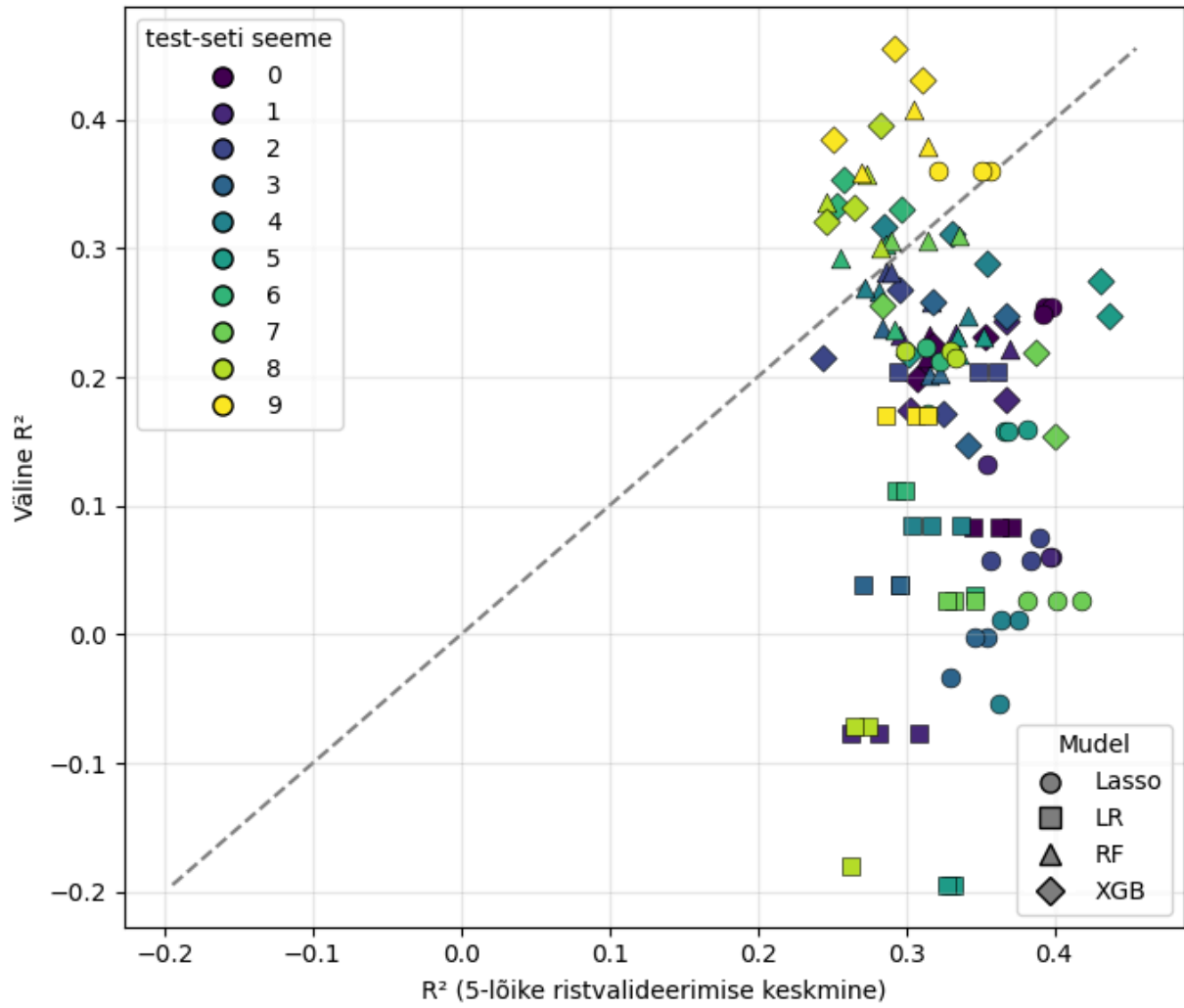
Joonis 3

Naiste keskmine väline R² väärtus vastavalt valitud tunnuste arvule



Joonis 4

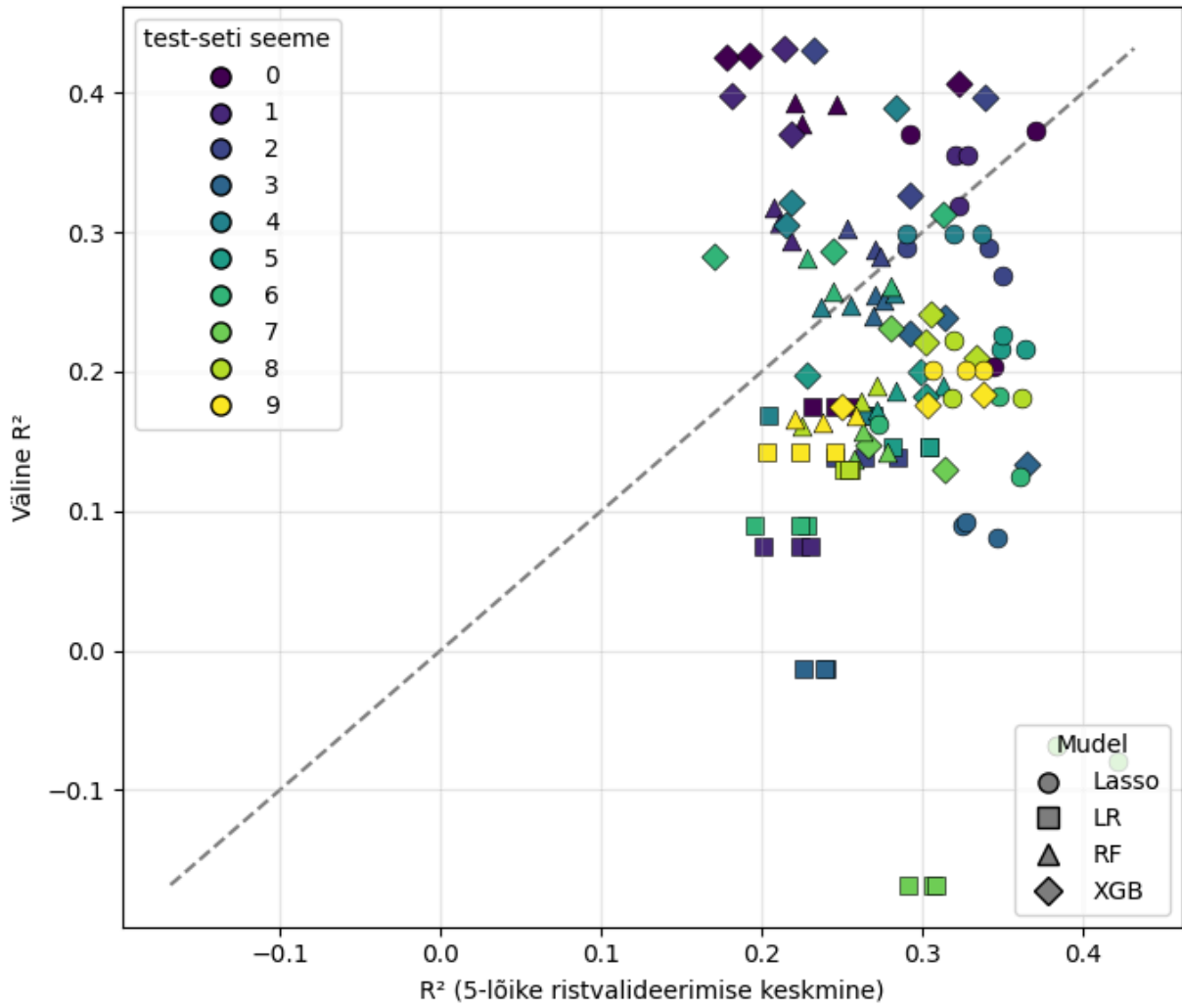
Meeste algne ja hilisem R^2 tulemus läbi nelja erineva mudeli erinevate seemnete kaupa



Märkus.

Joonis 5

Naiste algne ja hilisem R^2 tulemus läbi nelja erineva mudeli erinevate seemnete kaupa



Arutelu

See töö kasutas moodsaid masinõppe meetodeid, et püüa lüüa võimalikult lühike, aga hästi KMI-d ennustav küsimuste komplekt. Mudelite puhul esines mitmeid erinevusi KMI ennustatavuses nii stabiilsuse (standardhälve 0.115) kui ka tunnuste arvu osas. Kõige paremad tulemused olid puu-põhistel mudelitel. Nelja algoritmi kasutamisel oli kõige parem mudel XGBoost, mis oli stabiilselt tugev nii sisemise kui ka välise korrelatsiooniga ning mille vähendatud tunnused olid hea statistilise olulisusega KMI ennustamiseks. XGBoostiga leiti optimaalsete tunnuste kogum, mida lõpuks analüüsiti meeste ja naiste vahel. Tulemusena leiti meeste ja naiste ühisosad ning unikaalsed kogumid, mis seletavad soopõhiseid söömiskäitumisi.

Ennustusmudelite keskmine korrelatsioon on 0.191, meestel 0.178 ja naistel vastavalt 0.203. Parim mudel oli XGBoost, mille R^2 oli naistel keskmiselt 0.280 ja meestel 0.273. See tähendab, et naiste KMI on palju rohkem ennustatavam kui meeste oma. Samuti on tegemist arvestatava korrelatsiooniga, et ennustada KMI'd võrreldes terve andmestikuga.

Esmase treenimise järel oli kõige parem keskmine R^2 Lassol ($R^2=0.349$, võrdluseks XGBoost $R^2=0.295$), kuid pärast välise test-andmestiku sobitamist oli korrelatsioon $R^2=0.168$ (võrdluseks XGBoost $R^2=0.276$). See tähendab, et kuigi lineaarselt võib Lasso häid tulemusi anda, siis semantilist ja sügavamalt tähendust tunnuste vahel see nii hästi leida ei suuda. Samuti näitas head tulemust Otsustusmets (sisemine $R^2=0.279$, välimine $R^2=0.257$). Võib selgelt öelda, et parimad mudelid on antud tulemuste puhul otsustuspuude baasil olevad mudelid. Lineaarsetel mudelitel oli välise valiidsusega raskusi (RFECV mitmese regressiooni väline $R^2=0.060$, Lassol $R^2=0.168$).

Suur osa tulemustest oli üsna kõrge tunnuste arvuga (keskmine=149), millest suur osa oli tugeva R^2 väärtustega. Tundlikkus R^2 osas püsis mudeli tulemustes läbiv olenemata tunnuste arvust. XGBoost ja otsustuspuu tunnuste vahemikud olid 48 kuni 315. Lineaarsed mudelid jäid püsima madalate tunnuste arvu juures (mitmene regressioon $n\sim 15$, Lasso $n\sim 45$). Nende väline R^2 oli madalam ehk kokkuvõttes olid need ebastabiilsemad. Lineaarseid tunnuseid saaks veel peenhäälestada ja RFECV-le sööta nii, et tunnuseid oleks rohkem kui 15 või 45 ringis. Samuti vajaksid need mudeldamise loogika vahetust. Vastaval juhul üritavad lineaarsed mudelid saada deterministlikult lihtsa hea korrelatsiooni, mis võib sisemiselt olla tugev, kuid on väike seos välise test-andmestikuga ehk ei ole tugeva struktuuriga.

Tunnuseid vastavalt soole eksisteeris samu ja erinevaid, nii individuaalselt, kui ka küsimustike põhjal. Meeste läbivamad küsimused olid EPSI(n=17), DEBQ(n=14), EDEQ(n=8), IDAS-II(n=7) ja AUDIT(n=3) küsimustikest. Naiste läbivamad küsimused olid EPSI(n=14), EDEQ(n=7), IDAS-II(n=4) ja DERS(n=3) küsimustikest (Tabel 2). Söömishäirete Sümptomite Loendi (EPSI) küsimustik on oluline faktor nii meeste, kui ka naiste KMI ennustamisel. Naiste ja meestel ühiseid küsimusi antud tunnusega on 7. Näiteks selgitab see rahulolematust enda kehaga (nt epsi23, epsi24; vt Lisa C). Meeste puhul kerkivad esile küsimused "Ma arvan, et mu musklid on liiga väikesed" (epsi32) ja "Ma olen kasutanud proteiinilisandeid" (epsi35), mis seletavad maskuliinseid vajadusi ja perspektiive. Naiste puhul kerkisid keha rahulolematusega esile tunnused "Ma ei ole enda puusade suurusega rahul" (epsi34), "Olen kasutanud kaalulandusravimeid" (epsi17) ja "Kas sa oled tahtnud, et sa oleksid kõhn?" (edeq10). Tugeva mõjuga oli ka Söömishäirete Hindamise Küsimustik (EDEQ), mis sisaldas küsimusi nagu "Kui palju sa muretsed selle pärast, et teised inimesed näevad sind söömas?". Meeste läbiv küsimustik, mida naistel ei olnud, oli Hollandi Söömiskäitumise Küsimustik (DEBQ). See tähendab, et mehed olid söömisel rohkem mõjutatud negatiivsetest emotsioonidest. Samuti oli meestel statistiliselt piisavalt märkimisväärne ka Alkoholi Tarvitamise Häirete Tuvastamise Testi (AUDIT) esinemine, mis sisaldas tüüpilist alkoholi kuritarvitamist selgitavaid küsimusi. Üks stsenaarium võib olla, kus mehed tunnevad negatiivseid emotsioone ning kuritarvitavad stressi leevendamiseks alkoholi ja/või liigsöövad. Näiteks Hirvelä 2023 aasta teadustöös toodi välja, et mehed, kellel eksisteeris tugev tahe muskulaarsele kehale ning kes ei tundnud ennast piisavalt pikana, kuritarvitasid hilisemas eas rohkem alkoholi. Naiste puhul oli AUDIT'i ainuke küsimus: "Kui tihti viimase 12 kuu jooksul tundsid süütunnet pärast joomist?". Naistel tuli esile rohkem Emotsioonide Regulatsiooni Raskuste Skaalat (DERS), mis viitab impulsiivsele söömisele. Näiteks esines naiste puhul küsimusi "Ma tunnen liiga palju emotsioone ja ei suuda neid kontrollida" (ders3) ja "Kui ma olen pahane, siis ma võtan oma aja, et aru saada, mida ma päriselt tunnen" (ders34), samuti ka "Ma kaotasin kontrolli, mida ja kui palju ma sõin" (loc16). Meestel eksisteerisid samas vastupidiselt sõnastatud küsimused nagu "Ma pööran tähelepanu oma tunnetele" (ders6) ja "Ma tean täpselt kuidas ma ennast tunnen" (ders7), samas puudub neil täielik kontroll söömisharjumuste osas. See tähendab, et mehed on kohati liialt enesekindlad.

Tulemused olid volatiilsed ning sõltusid suuresti test-andmestiku seemnest, mis määras testandmete poolitamine juhuslikkuse. Seda näitab Joonis 1, kus osad tulemused on negatiivse

korrelatsiooniga ning osad tugeva positiivse korrelatsiooniga. See tähendab, et tõenäoliselt oli andmestik liiga väike, et stabiilseid tulemusi saada. Tulevikus oleks vaja suurendada andmestiku suurust, et stabiilsemaid mudeleid treenida ning ka seeläbi kindlamaid tunnuseid leida.

Soovitame suuremate treeningtsüklitega edasi uurida, kuidas ennustatavad tunnused omavahel kattuvad. Lineaarse regressiooni puhul võib proovida mahutada keskmise R^2 ühe standardhälbe sisse parimast R^2 -st. XGBoosti treenimise alla võiks veel rõhku pöörata, kuid samuti proovida tulemusi saada ka rohkemate tänapäevaste mudelite ja meetoditega. Lisaks tasub teha väga detailne peenhäälestamine kõikide mudelite peal, kuid tähelepanu tasub pöörata puu-põhiste mudelitele, mis leiavad erinevate küsimuste vahel paremaid semantilisi seoseid.

Kasutatud kirjandus

- Davis, C. (2013). From Passive Overeating to “Food Addiction”: A Spectrum of Compulsion and Severity. *ISRN Obesity*, 2013, 1–20.
<https://doi.org/10.1155/2013/435027>
- Racine, S. E., Hagan, K. E., & Schell, S. E. (2019). Is all nonhomeostatic eating the same? Examining the latent structure of nonhomeostatic eating processes in women and men. *Psychological Assessment*, 31(10), 1220–1233. <https://doi.org/10.1037/pas0000749>
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111.
<https://doi.org/10.1037/1040-3590.12.1.102>
- Vainik, U., Eun Han, J., Epel, E. S., Janet Tomiyama, A., Dagher, A., & Mason, A. E. (2019). Rapid Assessment of Reward-Related Eating: The RED-X5. *Obesity*, 27(2), 325–331.
<https://doi.org/10.1002/oby.22374>
- Vainik, U., García-García, I., & Dagher, A. (2019). Uncontrolled eating: a unifying heritable trait linked with obesity, overeating, personality and the brain. *European Journal of Neuroscience*, 50(3), 2430–2445. <https://doi.org/10.1111/ejn.14352>
- Vainik, U., Neseliler, S., Konstabel, K., Fellows, L. K., & Dagher, A. (2015). Eating traits questionnaires as a continuum of a single concept. Uncontrolled eating. *Appetite*, 90, 229–239. <https://doi.org/10.1016/j.appet.2015.03.004>
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning*, 46(1/3), 389–422.
<https://doi.org/10.1023/a:1012487302797>
- Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 1(1), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Priyatno, A. M., & Widiyaningtyas, T. (2024). A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 9(2), 196–207.
<https://doi.org/10.33480/jitk.v9i2.5015>

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/a:1010933404324>
- Sügis, E., Tampuu, A., Aljanaki, A., Fišel, M., & Kull, M. (2024). *Praktiline Andmeteadus*, Tartu Ülikooli arvutiteaduse instituut
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Tatjana van Strien, Jan, Gerard, & P.B. Defares. (1986). Dutch Eating Behaviour Questionnaire. *PsycTESTS Dataset*. <https://doi.org/10.1037/t04413-000>
- Forbush, K. T., Wildes, J. E., Pollack, L. O., Dunbar, D., Luo, J., Patterson, K., Petruzzi, L., Pollpeter, M., Miller, H., Stone, A., Bright, A., & Watson, D. (2013). Development and validation of the Eating Pathology Symptoms Inventory (EPSI). *Psychological assessment*, 25(3), 859–878. <https://doi.org/10.1037/a0032639>
- Latner, J. D., Mond, J. M., Kelly, M. C., Haynes, S. N., & Hay, P. J. (2014). The Loss of Control Over Eating Scale: development and psychometric evaluation. *The International journal of eating disorders*, 47(6), 647–659.
<https://doi.org/10.1002/eat.22296>
- Gormally, J., Black, S., Daston, S., & Rardin, D. (1982). The assessment of binge eating severity among obese persons. *Addictive behaviors*, 7(1), 47–55.
[https://doi.org/10.1016/0306-4603\(82\)90024-7](https://doi.org/10.1016/0306-4603(82)90024-7)
- Fairburn, C. G., & Beglin, S. J. (1994). Eating Disorder Examination Questionnaire. *PsycTESTS Dataset*. <https://doi.org/10.1037/t03974-000>
- Lowe, M. R., Butryn, M. L., Didie, E. R., Annunziato, R. A., Thomas, J. G., Crerand, C. E., Ochner, C. N., Coletta, M. C., Bellace, D., Wallaert, M., & Halford, J. (2009). The Power of Food Scale. A new measure of the psychological influence of the food environment. *Appetite*, 53(1), 114–118.
<https://doi.org/10.1016/j.appet.2009.05.016>

- Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., Stasik, S. M., & Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment, 19*(4), 399–420. <https://doi.org/10.1177/1073191112449857>
- Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction (Abingdon, England), 88*(6), 791–804. <https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>
- Bohn, K., Doll, H. A., Cooper, Z., O'Connor, M., Palmer, R. L., & Fairburn, C. G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour research and therapy, 46*(10), 1105–1110. <https://doi.org/10.1016/j.brat.2008.06.012>
- Gearhardt, A. N., Corbin, W. R., & Brownell, K. D. (2009). Preliminary validation of the Yale Food Addiction Scale. *Appetite, 52*(2), 430–436. <https://doi.org/10.1016/j.appet.2008.12.003>
- Hirvelä, L., Keski-Rahkonen, A., & Sipilä, P. N. (2023). Associations of broad eating disorder symptoms with later alcohol problems in Finnish adult twins: A nationwide 10-year follow-up. *The International Journal of Eating Disorders, 56*(10), 1854–1865. <https://doi.org/10.1002/eat.24013>

Lisa A**Meeste optimaalsete tunnuste küsimused kasutades mudelit XGBoost**

Järgnevad on küsimused formaadis "küsimustiku tunnuse nimi - küsimus". Listi esimene väärtus on osaleja vanus aastates.

1. RecodedAge - Vanus
2. audi12 - How often during the past 12 months did you become intoxicated or drunk from drinking any kind of...
3. audi2 - How many drinks of alcohol did you have on a typical day when you were drinking during the past 1...
4. audit3 - How often did you have 4 (for women) / 5 (for men) or more drinks on one occasion during the past...
5. bes1 - BES Below are groups of statements. Read all of the statements in each group and indicate the...
6. bes3 - -
7. cia14 - CIA INSTRUCTIONS Please choose the response option which best describes how your eating...-...made you feel like a failure?
8. debq15 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If you walk past the baker do you have the desire to buy something delicious?
9. debq20 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you get the desire to eat when you are anxious, worried or tense?
10. debq21 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you find it hard to resist eating delicious foods?
11. debq22 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you deliberately eat less in order not to become heavier?
12. debq26 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-How often do you try not to eat between meals because you are watching your weight?
13. debq28 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you have a desire to eat when you are bored or restless?

14. debq29 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-How often in the evening do you try not to eat because you are watching your weight?
15. debq3 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-Do you have a desire to eat when you have nothing to do?
16. debq33 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-When you are preparing a meal are you inclined to eat something?
17. debq4 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If you have put on weight, do you eat less than you usually do?
18. debq5 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-Do you have a desire to eat when you are depressed or discouraged?
19. debq6 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If food smells and looks good, do you eat more than usual?
20. debq7 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-How often do you refuse food or drink offered because you are concerned about your weight?
21. debq9 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If you see or smell something delicious, do you have a desire to eat it?
22. ders6 - DERS Please indicate how often the following statements apply to you by choosing the appropri...-I am attentive to my feelings.
23. ders7 - DERS Please indicate how often the following statements apply to you by choosing the appropri...-I know exactly how I am feeling.
24. edeq11 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Has thinking about shape or weight made it more difficult to concentrate on things you are interested in; for example, read, watch TV or follow a conversation?
25. edeq13 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you felt fat?
26. edeq14 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you had a strong desire to lose weight?

27. edeq15 - Over the PAST FOUR WEEKS (28 days), on what proportion of times that you have eaten have you felt...
28. edeq31 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How much would it upset you if you had to weigh yourself once a week for the next four weeks?
29. edeq32 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How dissatisfied have you felt about your weight?
30. edeq33 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How dissatisfied have you felt about your shape?
31. edeq36 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How uncomfortable have you felt about others seeing your body; for example, in a communal changing room, when swimming or wearing tight clothes?
32. epsi20 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I felt that overweight people are lazy
33. epsi22 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I planned my days around exercising.
34. epsi23 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I thought my butt was too big.
35. epsi24 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I did not like the size of my thighs.
36. epsi25 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I wished the shape of my body was different.
37. epsi30 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I felt that overweight people are unattractive
38. epsi32 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I thought my muscles were too small.
39. epsi35 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I used protein supplements.
40. epsi36 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-People encouraged me to eat more.

41. epsi4 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-People told me that I do not eat very much.
42. epsi40 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I tried to avoid foods with high calorie content
43. epsi41 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I exercised to the point of exhaustion.
44. epsi43 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I skipped two meals in a row.
45. epsi45 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I ate a very large amount of food in a short period of time (e.g., within 2 hours).
46. epsi6 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-People would be surprised if they knew how little I ate.
47. epsi7 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I used muscle building supplements.
48. epsi8 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I pushed myself extremely hard when I exercised.
49. idas18 - IDAS-II (© David Watson, 2011) Below is a list of feelings, sensations, problems, and experi...-I was worried about embarrassing myself socially
50. idas27 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I looked forward to things with enjoyment
51. idas48 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I found myself worrying all the time
52. idas8 - IDAS-II (© David Watson, 2011) Below is a list of feelings, sensations, problems, and experi...-I felt depressed
53. idas81 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I felt the urge to check to make sure I had done something
54. idas88 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I had so much energy it was hard for me to sit still
55. idas93 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I avoided talking about bad experiences from my past

56. pfs3 - PFS Please indicate the extent to which you agree that the following items describe you. Use the...-If I see or smell a food I like, I get a powerful urge to have some.

57. pfs8 - PFS Please indicate the extent to which you agree that the following items describe you. Use the...-Just before I taste a favorite food, I feel intense anticipation.

Lisa B**Naiste optimaalsete tunnuste küsimused kasutades mudelit XGBoost**

Järgnevad on küsimused formaadis "küsimustiku tunnuse nimi - küsimus". Listi esimene väärtus on osaleja vanus aastates.

1. RecodedAge - Vanus
2. audit7 - How often during the past 12 months did you have a feeling of guilt or remorse after drinking?
3. bes3 - -
4. bes8 - -
5. cia6 - CIA INSTRUCTIONS Please choose the response option which best describes how your eating...-...affected your ability to make everyday decisions?
6. debq22 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you deliberately eat less in order not to become heavier?
7. debq31 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you take into account your weight with what you eat?
8. debq4 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If you have put on weight, do you eat less than you usually do?
9. ders3 - DERS Please indicate how often the following statements apply to you by choosing the appropri...-I experience my emotions as overwhelming and out of control.
10. ders34 - Please indicate how often the following statements apply to you by choosing the appropriate numbe...-When I'm upset, I take time to figure out what I'm really feeling.
11. ders8 - DERS Please indicate how often the following statements apply to you by choosing the appropri...-I care about what I am feeling.
12. edeq10 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you definitely wanted your stomach to be flat?
13. edeq13 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you felt fat?
14. edeq14 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you had a strong desire to lose weight?

15. edeq31 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How much would it upset you if you had to weigh yourself once a week for the next four weeks?
16. edeq32 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How dissatisfied have you felt about your weight?
17. edeq34 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How concerned have you been about other people seeing you eat?
18. edeq7 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you been afraid of losing control over eating?
19. epsi1 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I did not like how my clothes fit the shape of my body.
20. epsi11 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I considered taking diuretics to lose weight.
21. epsi12 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I tried on different outfits, because I did not like how I looked.
22. epsi14 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I thought that obese people lack self-control
23. epsi17 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-I used diet pills.
24. epsi20 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I felt that overweight people are lazy
25. epsi22 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I planned my days around exercising.
26. epsi23 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I thought my butt was too big.
27. epsi24 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I did not like the size of my thighs.
28. epsi31 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I engaged in strenuous exercise at least 5 days per week.
29. epsi34 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I was not satisfied with the size of my hips

30. epsi36 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-People encouraged me to eat more.

31. epsi4 - EPSI© Below is a list of experiences and problems that people sometimes have. Read each item...-People told me that I do not eat very much.

32. epsi40 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I tried to avoid foods with high calorie content

33. idas47 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I found it difficult to talk with people I did not know well

34. idas49 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I had a very dry mouth

35. idas54 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I felt much worse in the morning than later in the day

36. idas78 - Below is a list of feelings, sensations, problems, and experiences that people sometimes have. Re...-I felt like I was ?on top of the world?

37. loc16 - LOCES In the last 4 weeks (28 days), how often have you had the following experiences duri...-I lost track of what and how much I was eating.

Lisa C

Naiste ja meeste optimaalsete tunnuste ühisosa küsimused kasutades mudelit XGBoost
Järgnevad on küsimused formaadis "küsimustiku tunnuse nimi - küsimus". Listi esimene väärtus on osaleja vanus aastates.

1. RecodedAge - None
2. bes3 - None
3. debq22 - Directions. For each item, decide if the item is true about you: never, rarely, sometimes, often...-Do you deliberately eat less in order not to become heavier?
4. debq4 - DEBQ Directions. For each item, decide if the item is true about you: never, rarely, sometimes, o...-If you have put on weight, do you eat less than you usually do?
5. edeq13 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you felt fat?
6. edeq14 - EDE-Q The following questions are concerned with the PAST FOUR WEEKS ONLY (28 days). Please r...-Have you had a strong desire to lose weight?
7. edeq31 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How much would it upset you if you had to weigh yourself once a week for the next four weeks?
8. edeq32 - OVER THE PAST FOUR WEEKS (28 DAYS) Please indicate the number which best describes your behavior.-How dissatisfied have you felt about your weight?
9. epsi20 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I felt that overweight people are lazy
10. epsi22 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I planned my days around exercising.
11. epsi23 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I thought my butt was too big.
12. epsi24 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-I did not like the size of my thighs.
13. epsi36 - Below is a list of experiences and problems that people sometimes have. Read each item to determi...-People encouraged me to eat more.

14. epsi4 - EPSI© Below is a list of experiences and problems that people sometimes have.

Read each item...-People told me that I do not eat very much.

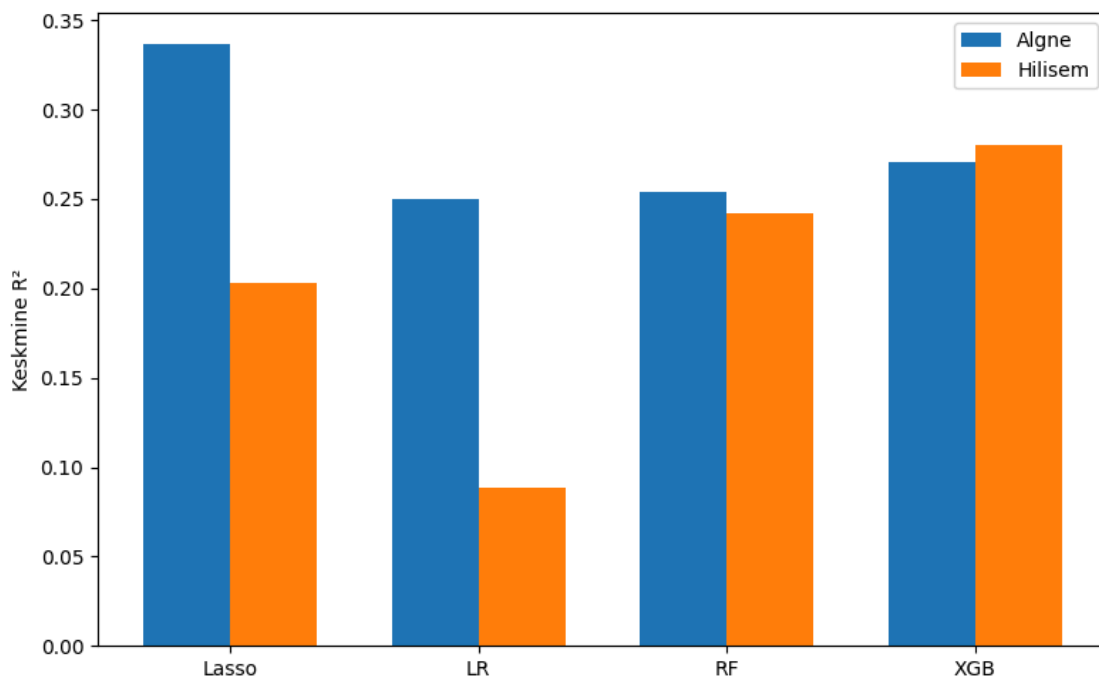
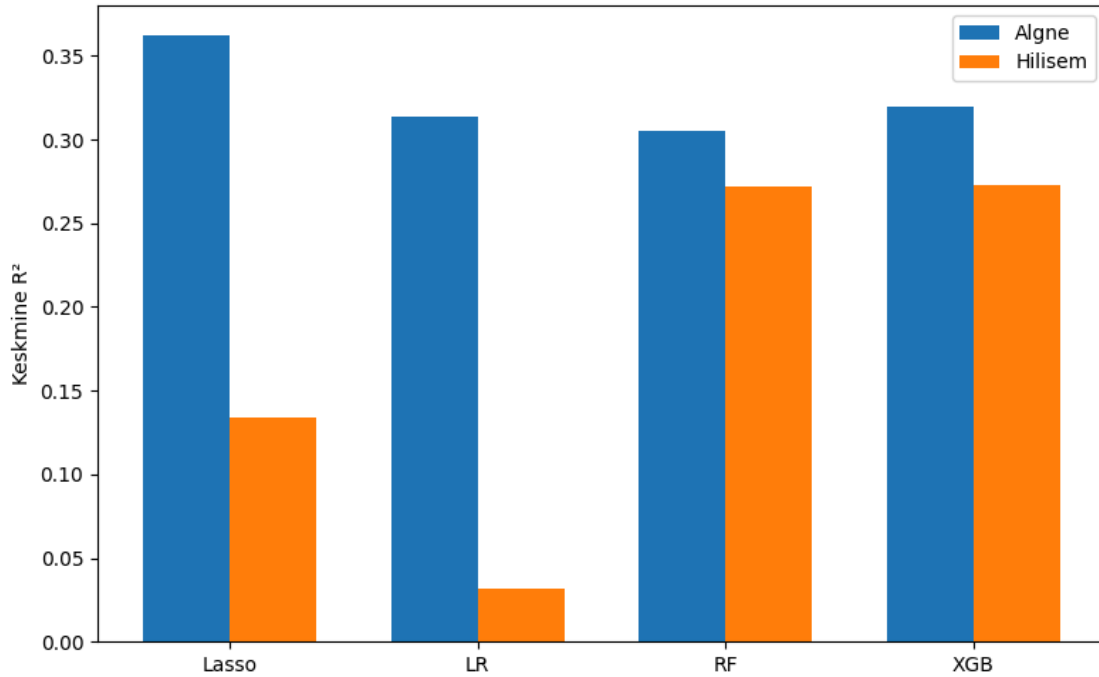
15. epsi40 - Below is a list of experiences and problems that people sometimes have. Read each

item to determi...-I tried to avoid foods with high calorie content

Lisa D

Sisemised ja välimised korrelatsioonid mudelite kaupa

Järgnevad on kaks tabelit visuaalseks ülevaateks leitud tunnuste valiidsustele. Kujutatud on esmalt nii meeste (ülemine), kui ka naiste varasemate ja hilisemate R^2 väärtusi.



Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.

Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis.

Karl Martin Kadaja