



# Betooni survetugevuse ennustamine kasutades andmekaevet

Bakalaureusetöö

Üliõpilane: Mark Dovydovich

Üliõpilaskood: 222531YAFB

Juhendajad: Dmitri Kartofelev, teadur

Küberneetika Instituut

Heiko J. Herrmann, vanemteadur

Küberneetik Instituut

Õppekava: Rakendusfüüsika

Tallinn 2026

# Autorideklaratsioon ja juhendaja resolutsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Mark Dovydovich, allkirjastatud digitaalselt.  
25.05.2026

Töö vastab bakalaureusetööle esitatavatele nõuetele.

Juhendaja: Dmitri Kartofelev, allkirjastatud digitaalselt.  
25.05.2026

# Tehisintellekti kasutamise deklaratsioon

Käesolevas töös on kasutatud tehisintellekti vaid koodi kirjutamisel ning koodivigade leidmisel.  
Töös on kasutatud ainult OpenAI ChatGPT-d.

Autor: Mark Dovydovich, allkirjastatud digitaalselt.

25.05.2026

# Sisukord

Sissejuhatus	2
1 Betoonisegu survetugevus	3
2 Kasutatud Pythoni teegi Scikit-Learn masinõppe mudelid	6
2.1 Scikit-learn teegi <i>Random Forest Regression</i> mudel . . . . .	6
2.2 Pythoni teegi Scikit-learn lineaarse regressiooni mudel . . . . .	9
3 Metoodika	10
4 Tulemused	14
5 Arutelu	22
Kokkuvõte	23
Tänuavaldused	24
Kasutatud kirjandus	25
A Lisad	29
A.1 Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks . . . . .	29
A.2 Link treenitud mudelite failidele. . . . .	30

# Sissejuhatus

Betoon on üks enimkasutatavaid ehitusmaterjale maailmas, mida kasutatakse laialt erineva otstarbega hoonete ja rajatiste ehitamisel. Betooni peamine kvaliteedinäitaja on selle survetugevus. Ehituses kasutatavad standardid näevad ette segude puhul kindlaid survetugevusi, mida segud peavad tardumisel saavutama 28 päeva jooksul. Euroopas reguleerib ehitusvaldkonda Eurokoodeks 2 [1]. Mida kõrgem on betoonisegu valmistamise kvaliteet seda suurem on selle lõplik survetugevus.

Töö eesmärgiks on treenida erinevaid masinõppe mudeleid ning võrrelda nende täpsust betoonisegude survetugevuse ennustamisel. Masinõppe mudelid peavad etteantud segu retsepti alusel hindama segu survetugevust sõltuvana ajast – ja seda täpsusega kuni 10%.

Varasemalt on taolisi uuringuid teinud A. Marchewka [2], kes on oma töös üritanud luua algoritme, et ennustada betooni survetugevust reaajas masinõppe abil jälgides betoonisegu hüdratsiooniprotsessi erinevate andurite abil. Samuti on teinud sarnast tööd ka M. Shaaban [3] keskendudes oma töös rohkem ANN ja DNN masinõppe mudelitele ning kõrgtugevatele betoonisegudele.

Betooni komponentide koostis on keeruline ning mittelineaarne protsess, mistõttu anda täpseid ja kvaliteetseid hinnanguid segu lõplikule survetugevusele osutub tihti keerukaks või pea võimatuks [2]. Betooni koostisse kuuluvad peamiselt 7 komponenti: tsement, räbu, tuhk, vesi, jämetäitematerjal (killustik), peentäitematerjal (liiv) ja plastifikaator. Betooni valmistamiseks segatakse komponendid teatud proportsioonides omavahel kokku ning jäetakse tarduma.

Kuigi on teada, et betoonisegu tardumisel on oluline roll ka temperatuuril, siis antud bakalaureusetöös kasutatud andmestikule ning standardite regulatsioonidele käsitletakse temperatuuri optimaalse ja konstantsena ning ei arvestata selle erinevaid mõjusid segude tardumisele.

Mudelite hindamisel vaadatakse olulisi mudelite meetrikuid ning kontrollitakse ennustuste täpsust koostöös Tallinna Tehnikaülikooli Ehituse ja Arhitektuuri Instituudi kolleegidega. Tulemuste kontrollimisel kasutatakse Rasmus Rahuoja magistritöö [4] raames tehtud katseteimikute andmeid.

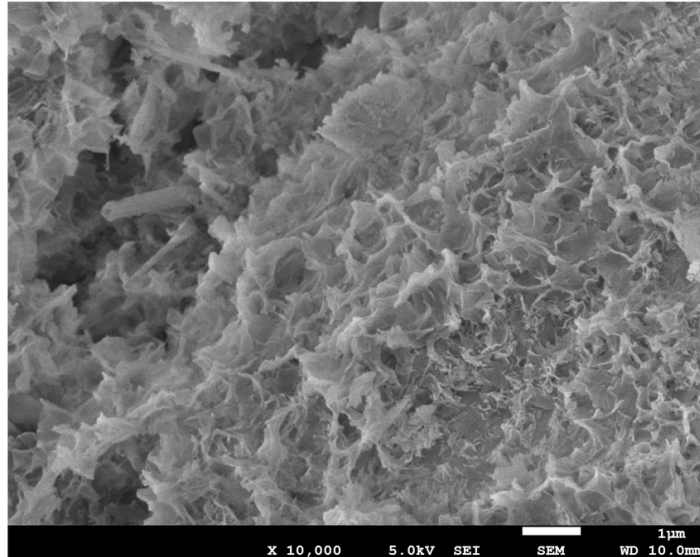
# 1. Betoonisegu survetugevus

Betoon on ehituses laialt levinud tehiskivim, mis saadakse tsemendi, vee ja täitematerjali segu tardumisel. Betoonisegus kasutatavad täitematerjalid ning nende suhe segus sõltuvad betooni tugevusklassist, millele segu peab tardudes vastama. Betooni survetugevust mõjutavad sajad tegurid alustades segu koostisest ja koostisosade proportsioonidest segus lõpetades keskkonna tingimuste ning nende muutlikkusega, milles betoonisegu tardub. Reaalses maailmas ei ole ehitusplatsil laboritingimuste tagamine võimalik. Selles töös keskendutakse vaid parameetritele, mis on kergesti kontrollitavad ning on osutunud kõige olulisemateks survetugevuse prognoosimisel: tsemendi, vee, räbu, tuha, peentäitematerjali, jämetäitematerjali ja plastifikaatori kogused segus ning tardumisperioodi pikkus [3, 5, 6]. Betoonisegu nõutav survetugevus oleneb ehitatava konstruktsiooni kasutusest.

Ohutuse ja kvaliteedikontrolli eesmärkidel on riiklikul tasandil kasutusel erinevate organisatsioonide ja ametiliitude poolt loodud standardid, mis määravad ära betoonisegu minimaalse lubatava survetugevuse ning ajavahemiku, mille jooksul see segu peab oma tulemuse saavutama. Väärtused on tingitud tsemendiklassidest ja tüüpidest ning nende valik toimub omakorda vastavalt konstruktsiooni otstarbele. Euroopa Liidu liikmesriikide territooriumil reguleerib betoonisegudega seonduvaid protsesse standard EN197-1 [7]. Iga Euroopa Liidu liikmesriik võtab omaks standardi, mis võib olla teatud määral kohandatud arvestades riigisiseseid eripärasid. Eesti Vabariigis on enamik betoonisegude olulisi parameetreid ning koostisosi reguleeritud standardi EVS-EN197-1:2011 poolt. [8]

Väga oluline mõju on betoonisegu hüdratsioonil. Kõige sagedamini kasutatav tsement, portlandtsement, koosneb peamiselt neljast (4) klinkri faasist: aliit ( $C_3S$ ), beliit ( $C_2S$ ), trikaltsiumaluminaat ( $C_3A$ ) ja kaltsium-aluminoferriit ( $C_4AF$ ). Aliit ehk trikaltsiumsilikaat on neljast eespool mainitud klinkri faasist kõige reaktiivsem ning mõjutab oluliselt betoonisegu survetugevuset segu tardumise algfaasis. Beliit (dikaltsiumsilikaat) reageerib aliidist aeglasemalt, kui mängib olulist rolli betoonisegu pikaajalises survetugevuse arengus. Trikaltsiumaluminaat reageerib veega eksotermiliselt ja väga kiiresti. Kiire reageerimine võib viia betooni vältardumiseni. Vältardumine on fenomen betoonikeemias, mille korral betoon tardub väga kiiresti, mistõttu tema edasine töötlemine on raskendatud. Selle vältimiseks lisatakse tsemendile kipsi. Aluminaat-ferriit reageerib veega sarnaselt aliidile, kuid mängib olulist rolli betoonisegu pikaajalisel vastupidavusel ja stabiilsuse säilimisel. [2]

Betoonisegu tardumise faaside kiirus oleneb suuresti temperatuurist, mistõttu betoonisegu tardumise algfaasis on vaja tagada keskkonna kindel niiskus ja temperatuur ( $20\text{ }^\circ\text{C}$ ) [2]. Kõrgematel temperatuuridel tardub segu kiiremini ning on suurema tihedusega, väiksema poorsusega. Segu jõuab soovitud survetugevuseni küll kiiremini, kuid kuivemate tardumistingimuste tõttu suureneb mõranemise oht. Betoonisegu hoitakse tardumisperioodi alguses (vähemalt 7 päeva, ideaalis 28 päeva) täielikult vee all, et tagada segu ühtlane ja pidev reaktsioon veega ning



Joonis 1.1: C-S-H peenstruktuur. Allikas: [9].

vältida kuivtardumist [6].

Räbu ja tuhk on pozzolaansed<sup>1</sup> materjalid, mille peamine põhjus segusse lisamiseks on võime tekitada kaltsiumsilikaathüdraati<sup>2</sup> (C-S-H). Õige tuha kogus võib tõsta survetugevust langetades tardumisperioodi alguses segu temperatuuri, mille tõttu vajab segu vähem vett. Samuti aitab tuha kasutus suurendada C-S-H teket segus, mis samuti langetab tsemendi poorsust ning tihendab segu mõjutades positiivselt survetugevust. Siiski suure tuha koguse juures saavutab betoon ettenähtud perioodi jooksul väiksema survetugevuse, sest tuhk mõjutab settimise kiirust läbi segu temperatuuri languse. [6] C-S-H peamine funktsioon on segu täitematerjali sidumine ning poorsuse vähendamine, mille kaudu omandab betoon oma karakteristikliku survetugevuse. C-S-H geeli poorne struktuur on nähtav vaid mikroskoobi abil (vt Joonis 1.1). Arvestades, et peen- ja jämetäitematerjalide terade suurus on tüüpiliselt mm-skaalas, on C-S-H geel ideaalne lahendus betooniseгу sideaine poorsuse vähendamiseks ning survetugevuse tõstmiseks, sest see suudab täita vabat ruumi täitematerjalide terade vahel tagades segu hermeetilisuse. [6]

Jämetäitematerjal (killustik) mängib olulist rolli betooni survetugevuse ja jäikuse tõstmises. Killustiku iga tera peab purunemata ja pragunemata võtma vastu konstruktsiooni koormust ning jaotama selle üle naaberterade mehaanilise kontakti kaudu. Jämetäitematerjali halva valiku korral on betooniseгу survetugevus ja jäikus piiratud täitematerjali purunemisjäikuse poolt. [6]

Peentäitematerjali (liiva) peamised ülesanded on killustiku terade vahel oleva tühja ruumi täitmine, segu tiheduse suurendamine ning betooniseгу töödeldavuse parendamine. Täites ära suuremat osa ruumist killustiku terade vahel vähendab liiv tsemendi ja vee lisamise vajadust. Väga peeneteraline liiv täidab ära rohkem tühja ruumi killustiku terade vahel, kuid suurendab vajaliku vee kogust. [6]

<sup>1</sup>Materjal, mis reageerib vee ja kaltsiumhüdroksiidiga moodustades kaltsiumsilikaathüdraadi (C-S-H) [6]

<sup>2</sup>Pozzolaansete materjalide reageerimisel vee ja  $\text{Ca}(\text{OH})_2$ -ga tekkiv geelitaoline aine

Plastifikaatori peamine ülesanne on betoonisegus kasutatava vee koguse vähendamine. Plastifikaatorid jaotavad osakesi segus elektrostaatiliste ja steeriliste efektide abil. Kuna vee koguse suurendamine mõjutab negatiivselt betooni survetugevust, aitab plastifikaatorite (tänapäeval superplastifikaatorite) kasutamine vähendada vee kogust segus, mis võib tõsta betoonisegu survetugevust. Liigne plastifikaatori kogus segus võib põhjustada aeglast tardumist ja roomet. Samuti raskendab see segu homogeniseerumist. Kõik need asjaolud põhjustavad betoonisegu survetugevuse vähenemist ning suurendavad konstruktsiooni purunemise riski. [6, 10]

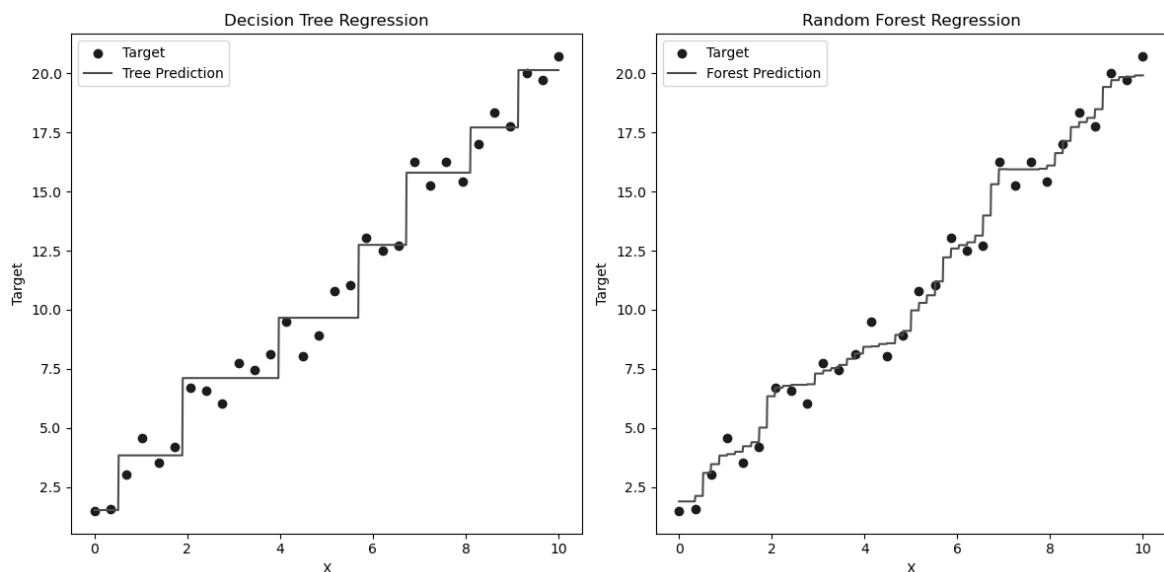
Betooni survetugevuse kasv on kõige suurem segu tardumisperioodi alguses. Betoon saavutab oma kõige tugevama oleku ligi 25 aasta pärast, mille järel hakkab segu survetugevus langema. [11]

## 2. Kasutatud Pythoni teegi Scikit-Learn masinõppe mudelid

Betoonisegu komponentide keeruka koosmõju tõttu on selle survetugevuse kasv mittelineaarne [11, 12]. Selmet treenida andmekogul kõiki Scikit-learn mudelid (k.a lineaarseid) tuli valida mudelid, mille lähendusalgoritm suudavad tuvastada ka mittelineaarseid seoseid. Scikit-learn teek pakub kasutamiseks mitmeid mittelineaarseid masinõppe mudelid: otsustuspuudel põhinevad mudelid, ANN (*Artificial Neural Network*) mudelid, näitepõhise õppega mudelid (nt *k-nearest neighbors*) jt. Käesoleva töö raames valiti otsustuspuu-põhine juhusliku metsa regressiooni masinõppe mudel (ingl. k. *Random Forest Regression model*) andmekogu lihtsus- tuste (8 sisendparameetrit: 7 koostisosa ja tardumisperioodi pikkus ning 1 väljundparameeter – survetugevus) ning mudeli intuiitsuse tõttu. Võrdluseks toodi välja ka üksiku otsustuspuu regressiooni ning lineaarse regressiooni mudelite ennustusi. Kõige põhjalikumalt, siiski, on käsitletud juhusliku metsa regressiooni mudeli ennustusi ning neile eelnevat mudeli treenimist.

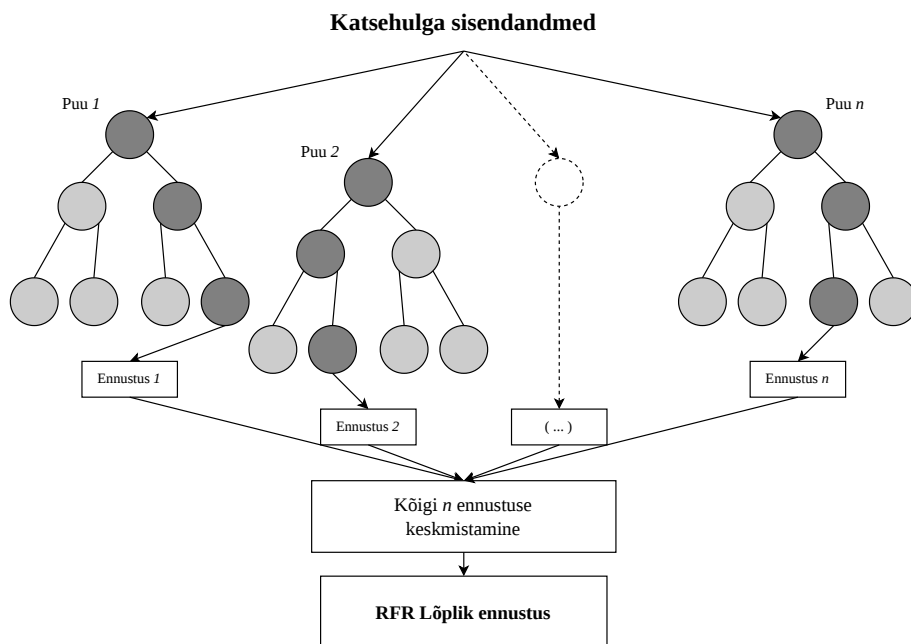
### 2.1. Scikit-learn teegi *Random Forest Regression* mudel

Juhusliku metsa regressiooni (RFR) ja otsustuspuu regressiooni (DTR) mudelid on niimetatud *step*-tüüpi mittelineaarse regressiooni masinõppe mudelid, mille iseloomulikuks tunnuseks on graafikul tekkivad ennustuste “astmed” (vt Joonis 2.1). [5, 13, 14]



Joonis 2.1: DTR ja RFR mudelite astmelisus parameetri  $x$  ennustamisel. Allikas: [14].

RFR saavutab tihedama sammu tänu mitmete üksikute “puude” treenimisele ning nende



Joonis 2.2: RFR mudeli tööpõhimõte.

keskmistatud väärtuse arvutamisele. RFR on statistiline mudel, mis keskmistab üksikute puude väärtused, et tagada tulemusi täpsemini prognoosiv ennustuskõver (vt Joonis 2.1).

Mudeli treenimise korral jagatakse olemasolev andmekogu kaheks osaks: treeningandmed ja testandmed. Testandmeid kasutatakse treeningprotsessi järgselt mudeli valideerimiseks. Andmekogu jagamise peamine eesmärk on valideerida mudel andmetel, mida mudel pole varem "näinud". See kindlustab, et arvutatav mudeli viga ei tule liiga optimistlik selle tõttu, et ennustusi tehakse andmetel, mille tegelikud väärtused on mudelile juba teada. Tavaliselt jagatakse andmekogu nii, et 80% andmetest moodustaks treeninguks kasutatavad andmed ning 20% moodustaks testandmed. Treeningandmete korral on tegu andmekogu osaga, mille põhjal hakkab mudel otsima korrelatsiooni sisendparameetrite vahel. [15]

Otsustuspuude loomisel ja mudelite treenimisel tuleb võtta arvesse mitmeid olulisi treeningparameetreid nagu üksiku puu sügavus (int: `max_depth`), puude arv mudelis (int: `n_estimators`), andmehulga üksikutest ridadest koosneva hulga minimaalne suurus (int: `min_sample_split`), viimase "lehe" minimaalne ridade arv, mille põhjal otsustus tehakse ning maksimaalne muutujate arv (`max_features`), mis piirab seda, mitut komponenti andmekogu ühest reast 1 treenitav puu tohib korraga arvesse võtta. Mudeli viga minimeeritakse vähimruutude meetodi abil. Vähimruutude meetodi korral eesmärgiks on minimeerida summat

$$Q = \sum_{i=1}^n [y_i - (\vec{z} \cdot \vec{\beta})_i]^2, \quad (2.1)$$

kus  $n$  on mudeli puude arv, kordajad  $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_8)$ , mis on iga koostisosa ja tardumisaja jaoks mudeli treenimisel saadud kordajad,  $y_i$  on katseliselt määratud survetugevus ning  $z_i$  on sisendväärtuste vektor. Mudelis on kordajad  $\vec{\beta}$  erineva väärtusega iga puu lõikes.

Mudeli pikaajalise treenimise korral vahe ruut kahaneb iga iteratsiooniga. Siiski, iteratsiooni arvu kasvades suureneb treenimise aeg ning väheneb treeningu efektiivsus. See tähendab, et iga järgmine treeningu iteratsioon annab väiksema paranduse mudeli ennustustäpsusele kui eelnev iteratsioon. Lisaks suureneb mudeli ületreenimise oht, mis tähendab, et mudel suudab prognoosida aina täpsemini andmeid, mille peal seda treeniti, kuid selle arvelt väheneb mudeli täpsus andmete korral, mida mudel pole veel “näinud”. Treenimise optimeerimiseks on võimalik kasutada erinevaid meetodeid ning nende kombinatsioone. Selleks, et vältida mudeli ületreenimist (ingl. k. *overfitting*) on võimalik kasutada GridSearch algoritmi. GridSearch algoritm võimaldab kasutajal automatiseerida erinevate hüperparemetrite alusel mudelite treenimist aktsepteerides sisendväljadesse mudeli hüperparameetrite loendeid. [13, 16, 17]

Kõige levinumate meetodite nimed mudeli treenimisel on AdaBoost, Bagging, XGBoost, ristkontroll jt. AdaBoost (*Adaptive Boosting*) ja XGBoost (*Extreme Gradient Boosting*) on võimendusmeetodid, mille korral toimub mudeli loomine iteratiivselt. Iga järgnev iteratsioon keskendub eelneva treeningu käigus tehtud vigade parandamisele. AdaBoost seab igale valesti ennustatud väärtusele kaalu vastavalt sellele kui suure vea mudel selle ennustamisel tegi. Järgmise iteratsiooniga keskendutakse rohkem nendel väärtustel, mille prognoosimisel tehti suurema kaaluga vead. XGBoost rakendab iteratiivset vea parandust gradientvõimenduse abil. Gradientvõimenduse korral ei omistata üksikutele väärtustele eraldi kaalusid, vaid vaadatakse mis suunas liikudes on vea vähenemine kiireim. Vea vähendamise tõhusust hinnatakse ruutkeskmise vea abil (2.2). [17]

$$\text{MSE}(\hat{y}_i) = \sum (y_i - \hat{y}_i)^2, \quad (2.2)$$

kus  $y_i$  on tegelik survetugevus mõõtevea piires ning  $\hat{y}_i$  on mudeli ennustus ettenatud koostise alusel. Bagging (*Bootstrap aggregating*) on masinõppe mudeli treenimise meetod, kus iga mudel treenitakse andmekogu juhuslikult valitud alamandmehulga peal ning seejärel erinevate mudelite hinnangud keskmistatakse. Bagging’u algoritmi kasutatakse mudeli treenimisel vaikimisi, kui kodeerija ei määra teisiti. Erinevate parameetrite täpne kirjeldus tuleneb otsustuspuude puudulikkusest, millega välditakse ületreenimist. Mitmete ebatäiuslike puude keskmistatud väärtus annab täpsema hinnangu kui ühe väga hästi treenitud otsustuspuu hinnang, sest ebatäiuslike puude korral ei pea iga puu suutma kirjeldada tervet andmekogu, vaid teatud osa sellest. Kogu metsa kaasamisel igal ennustusel suudab mudel ennustada täpsemalt ka erijuhte (vt Joonis 2.2). [13]

## 2.2. Pythoni teegi Scikit-learn lineaarse regressiooni mudel

LR mudel töötab tavapärase vähimruutude meetodi (2.1) põhjal. Mudeli algoritm otsib andmete kirjeldamiseks sirget, mille kõrvalekalle kõikide punktide suhtes on vähim võimalikest. Lineaarsed mudelid sobivad hästi kirjeldamiseks olukordi, kus vaadeldakse üksteisest sõltuvaid parameetreid paaritsi, kuid suurema sõltuvate muutujate arvuga muutub mudel kiiresti ebatäpseks. Samuti ei ole mõttekas kasutada lineaarset regressiooni olukordades, kus parameetrite vahel puudub lineaarne seos. Betooni survetugevuse ennustamise korral on tegu vähemalt 9 parameetriga, mis sõltuvad teineteisest mittelineaarselt. Samuti on näha, et teatud parameetrite omavaheline seos on väga nõrk või puudub (vt Joonised 4.2, 4.5, 4.6).

Samuti ei ole mitmekordne lineaarne regressioon (2.3) antud probleemi lahendamisel eelistatud meetod. Kuigi segus kasutatavate teatud komponentide proportsioonid sõltuvad lineaarselt (nt mida rohkem pannakse peentäitematerjali, seda vähem saab panna jämetäitematerjali), siis survetugevusele avaldub nende koosmõju mittelineaarselt. Seose mittelineaarsust kinnitavad ka varasemad uuringud [2, 3, 5].

Mitmekordne lineaarne regressioon on esitatav funktsiooni

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2.3)$$

näol, kus otsitav parameeter  $y$  on avaldatav sisendparameetrite  $x_i$  lineaarse kombinatsioonina vastavate kaaludega  $\beta_i$  ja prognoosijäägiga  $\varepsilon$ . [16]

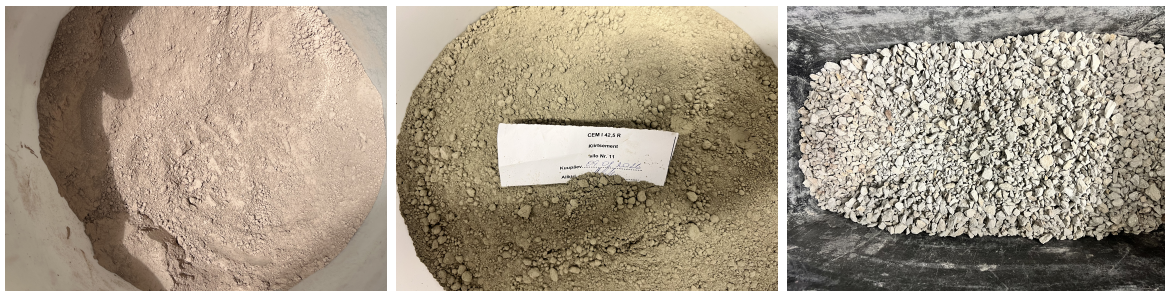
### 3. Metoodika

Töö raames treeniti 3 mudelit ning võrreldi mudelitäpsusi kasutades UCI (*University of California, Irvine*) masinõppe repositooriumis saadaval olevat andmekogu [18]. Mudelite loomiseks ja treenimiseks kasutati programmeerimiskeeles Python™ loodud teeki SciKit-Learn. Andmekogus on esitatud 6 põhilist segu koostiosa, tardumisperiodid päevades ning katseliselt mõõdetud survetugevus (vt tabel 3.1). Andmekogu sisaldab 1031 segu andmeid, mille valimisel oli andmekogu koostaja valinud betoonisegude koostiseid nii, et jätta välja tootjapoolseid standardiväliseid tootmistehnoloogiaid. [5, 18]

Tabel 3.1: Mudeli treenimiseks kasutatud andmekogu segment.

Tsement	Räbu	Tuhk	Vesi	Plastifikaator	Killustik	Liiv	Vanus	$\sigma_N$
$\frac{\text{kg}}{\text{m}^3}$							päev	MPa
540	0	0	162	2.5	1040	676	28	79,99
540	0	0	162	2.5	1055	676	28	61,89
332.5	142.5	0	228	0	932	594	270	40,27

Andmekogus kasutati Portlandi I-tüüpi tsementi, mille täpsem koostis on kirjeldatud ASTM C150 standardis [19] leitavas tabelis, mille väljavõte on toodud allpool (vt Joonis 3.1). Andmekogu autori väitel on kasutatud räbu pärit Taiwani metallitööstusest.



(a) Põlevkivituhk

(b) Tsement

(c) Jämetäitematerjal

Joonis 3.2: Magistritöö [4] segus kasutatud põlevkivituhk, tsement ja killustik.

Tuhk on pärit Taiwani elektrijaamast. Veena kasutati tavalist kraanivett, mille sisaldus vastab Taiwani kraanivee puhtust reguleerivatele standarditele. Kasutatud killustik ja liiv on samuti Taiwani päritoluga. Killustiku jaoks kasutati Taiwani purustatud looduslikku kivimit, mille terade läbimõõt ei ületa 10 mm. Liivana kasutati pestud looduslikku Taiwani jõgedest kaevatavat liiva, mille terad on peensusmooduliga 3,0. Sellist täitematerjali loetakse suhteliselt jämedaks ning valdav osa terade läbimõõte jääb harilikult vahemikku 0,6–1,2 mm. Plastifikaatorina kasutati vastavalt ASTM C494 standardile G-tüüpi naaftaleenformaldehüüdi kondensaati ja

TABLE 1 Standard Composition Requirements						
Cement Type <sup>4</sup>	Applicable Test Method	I and IA	II and IIA	III and IIIA	IV	V
Aluminum oxide (Al <sub>2</sub> O <sub>3</sub> ), max, %	C 114	...	6.0	...	...	...
Ferric oxide (Fe <sub>2</sub> O <sub>3</sub> ), max, %	C 114	...	6.0 <sup>B,C</sup>	...	6.5	...
Magnesium oxide (MgO), max, %	C 114	6.0	6.0	6.0	6.0	6.0
Sulfur trioxide (SO <sub>3</sub> ), <sup>D</sup> max, %	C 114					
When (C <sub>3</sub> A) <sup>E</sup> is 8 % or less		3.0	3.0	3.5	2.3	2.3
When (C <sub>3</sub> A) <sup>E</sup> is more than 8 %		3.5	<sup>F</sup>	4.5	<sup>F</sup>	<sup>F</sup>
Loss on ignition, max, %	C 114	3.0	3.0	3.0	2.5	3.0
Insoluble residue, max, %	C 114	0.75	0.75	0.75	0.75	0.75
Tricalcium silicate (C <sub>3</sub> S) <sup>E</sup> , max, %	See Annex A1	...	...	...	35 <sup>B</sup>	...
Dicalcium silicate (C <sub>2</sub> S) <sup>E</sup> , min, %	See Annex A1	...	...	...	40 <sup>B</sup>	...
Tricalcium aluminate (C <sub>3</sub> A) <sup>E</sup> , max, %	See Annex A1	...	8	15	7 <sup>B</sup>	5 <sup>C</sup>
Sum of C <sub>3</sub> S + 4.75C <sub>3</sub> A <sup>G</sup> , max, %	See Annex A1	...	100 <sup>H</sup>	...	...	...
Tetracalcium aluminoferrite plus twice the tricalcium aluminate (C <sub>4</sub> AF + 2(C <sub>3</sub> A)), or solid solution (C <sub>4</sub> AF + C <sub>2</sub> F), as applicable, max, %	See Annex A1	...	...	...	...	25 <sup>C</sup>

<sup>A</sup>See Note 2.  
<sup>B</sup>Does not apply when the heat of hydration limit in Table 4 is specified.  
<sup>C</sup>Does not apply when the sulfate resistance limit in Table 4 is specified.  
<sup>D</sup>There are cases where optimum SO<sub>3</sub> (using Test Method C 563) for a particular cement is close to or in excess of the limit in this specification. In such cases where properties of a cement can be improved by exceeding the SO<sub>3</sub> limits stated in this table, it is permissible to exceed the values in the table, provided it has been demonstrated by Test Method C 1038 that the cement with the increased SO<sub>3</sub> will not develop expansion in water exceeding 0.020 % at 14 days. When the manufacturer supplies cement under this provision, he shall, upon request, supply supporting data to the purchaser.  
<sup>E</sup>See Annex A1 for calculation.  
<sup>F</sup>Not applicable.  
<sup>G</sup>See Note 4.  
<sup>H</sup>In addition, 7-day heat of hydration testing by Test Method C 186 shall be conducted at least once every six months. Such testing shall not be used for acceptance or rejection of the cement, but results shall be reported for informational purposes.

Joonis 3.1: Tsemendi koostis tüübi järgi vastavalt ASTM C150 standardile. Allikas: [19]

aminohappe kopolümeeri sisaldavat plastifikaatorit. [5, 20]

GridSearch algoritmiga treeniti 12500 erinevat mudelit, mille seast valis algoritm mudeli meetrikute alusel parima. Mudeli treenimisel kasutati tavapäraselt *Bagging* algoritmi, mille korral andmed jagatakse alamhulkadeks ning seejärel kasutatakse neid treenimiseks. Treenimiseks kasutati 80% andmekogust, seejärel mudeli valideerimiseks 20%. Mudelid hinnati 5-kordse ristkontrolli abil, mille eesmärgiks oli juhuslikult tekkivate “soodsate” treeningandmete jaotuste vältimine. 5-kordne ristkontroll aitab valida treenitud RFR mudelite hulgast parima. Ristkontrolli käigus tagastab funktsioon iga  $k$  valimi kohta täpssuskoorid, mille seast valitakse parima täpssuskooriga mudel [13].

Tabel 3.2: Kasutatud treeningparameetrite võrgustikud. GridSearch algoritmi raames tähendab hüperparameetri väärtus `None`, et kasutaja ei sea omapoolset piirangut hüperparameetri valikule ning algoritm kasutab vaikimisi seatud väärtust [21].

<i>Random Forest</i>	<i>Decision Tree</i>
n_estimators: [100, 150, 200, 250, 300]	–
max_depth: [None, 5, 10, 15, 20]	max_depth: [None, 5, 10, 15, 20]
min_samples_split: [2, 5, 10, 15, 20]	min_samples_split: [2, 5, 10, 15, 20]
min_samples_leaf: [1, 2, 4, 8, 16]	min_samples_leaf: [1, 2, 4, 8, 16]
max_features: [sqrt, log2, 0.5, None]	max_features: [sqrt, log2, 0.5, None]

Treenimisel rakendati RFR ja DTR mudelitele treeningparameetrite võrkusid (vt Tabel 3.2), et suurendada “hästi” treenitud mudelite osakaalu kõikide seas. Kõikide mudelite korral tagastati ka nende keskmised ruutvead, keskmised absoluutvead ja nende determinatsioonikordajate ruutväärtused  $R^2$ . Lineaarse regressiooni mudelile vastavat parameetrite võrgustikku rakendada ei ole võimalik selle tööprintsibi tõttu.

TABLE 3 Standard Physical Requirements									
Cement Type <sup>A</sup>	Applicable Test Method	I	IA	II	IIA	III	IIIA	IV	V
Air content of mortar, <sup>B</sup> volume %:	C 185								
max		12	22	12	22	12	22	12	12
min		...	16	...	16	...	16	...	...
Fineness, <sup>C</sup> specific surface, m <sup>2</sup> /kg (alternative methods):									
Turbidimeter test	C 115								
Average value, min <sup>D</sup>		160	160	160	160	...	...	160	160
Any one sample, min <sup>E</sup>		150	150	150	150	...	...	150	150
Average value, max <sup>D</sup>		...	...	240 <sup>F</sup>	240 <sup>F</sup>	...	...	240	...
Any one sample, max <sup>E</sup>		...	...	245 <sup>F</sup>	245 <sup>F</sup>	...	...	245	...
Air permeability test	C 204								
Average value, min <sup>D</sup>		280	280	280	280	...	...	280	280
Any one sample, min <sup>E</sup>		260	260	260	260	...	...	260	260
Average value, max <sup>D</sup>		...	...	420 <sup>F</sup>	420 <sup>F</sup>	...	...	420	...
Any one sample, max <sup>E</sup>		...	...	430 <sup>F</sup>	430 <sup>F</sup>	...	...	430	...
Autoclave expansion, max, %	C 151	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
Strength, not less than the values shown for the ages indicated as follows: <sup>G</sup>									
Compressive strength, MPa (psi):	C 109/ C 109M								
1 day		...	...	...	...	12.0 (1740)	10.0 (1450)	...	...
3 days		12.0 (1740)	10.0 (1450)	10.0 (1450) 7.0 <sup>H</sup> (1020) <sup>H</sup>	8.0 (1160) 6.0 <sup>H</sup> (870) <sup>H</sup>	24.0 (3480)	19.0 (2760)	...	8.0 (1160)
7 days		19.0 (2760)	16.0 (2320)	17.0 (2470) 12.0 <sup>H</sup> (1740) <sup>H</sup>	14.0 (2030) 9.0 <sup>H</sup> (1310) <sup>H</sup>	...	...	7.0 (1020)	15.0 (2180)
28 days		...	...	...	...	...	...	17.0 (2470)	21.0 (3050)
Time of setting; Vicat test: <sup>I</sup>	C 191								
Time of setting, min, not less than		45	45	45	45	45	45	45	45
Time of setting, min, not more than		375	375	375	375	375	375	375	375

<sup>A</sup>See Note 2.  
<sup>B</sup>Compliance with the requirements of this specification does not necessarily ensure that the desired air content will be obtained in concrete.  
<sup>C</sup>The testing laboratory shall select the fineness method to be used. However, when the sample fails to meet the requirements of the air-permeability test, the turbidimeter test shall be used, and the requirements in this table for the turbidimetric method shall govern.  
<sup>D</sup>Average value shall be determined on the last consecutive five samples from a source.  
<sup>E</sup>The value of any one sample shall be the result of a test or average of tests on any one sample.  
<sup>F</sup>Maximum average and maximum single sample fineness limits do not apply if the sum of C<sub>3</sub>S + 4.75C<sub>3</sub>A is less than or equal to 90.  
<sup>G</sup>The strength at any specified test age shall be not less than that attained at any previous specified test age.  
<sup>H</sup>When the optional heat of hydration in Table 4 is specified.  
<sup>I</sup>The time of setting is that described as initial setting time in Test Method C 191.

Joonis 3.3: Tsemendi survetugevus 1, 3, 7 ja 28 päeva järel vastavalt ASTM C150 standardile. Allikas: [19].

Edaspidi kontrolliti parimat mudelit koostöös Tallinna Tehnikaülikooli Ehituse ja Arhitektuuri Instituudiga. Betoonisegused, mida kontrollimiseks kasutati, valmistati magistritöö [4] raames. Kõik täpsemad arvutusmeetodid, millega pandi paika segukomponentide proportsioonid segus, on kirjeldatud vastavas magistritöös. Betoonsegu valmistamisel kasutati toatemperatuuril joogivett vastavalt standardile EVS-EN 1008:2002 [22]. Tsemendina kasutati CEM I 42,5 R kiirtsementi. CEM I viitab, et tegu on portlandtsemendiga, mis koosneb 96% ulatuses jahvatatud klinkrist koos kipsiga. 4% moodustavad lisakoostisosad, mis on täpsemalt mainitud vastavas magistritöös. Kasutatud põlevkivituhk on pärit Balti elektri jaama 8. ploki filtrist ning on töödeldud vajalikke sideaineliste omaduste saavutamiseks (vt Joonis 3.2a).

Superplastifikaator on vesialuselistel akrüülpolümeeridel põhinev plastifikaator, mis tagab osakeste parema jaotumise segus. Jämetäitematerjalina kasutati Väo karjäärist kaevandatud ning Limestone factories of Estonia OÜ poolt töödeldud paekivikillustikku. Peentäitematerjalina kasutati Männiku karjäärist kaevatud "M2-EL" tähisega jämedat ehitusliiva fraktsiooniga 0-4 mm, mille tootmise eest vastutas ettevõtte AS Silikaat. [4]

Oluline on märkida, et andmekogus [18] kasutatud portlandtsement peab 28 päeva järel

saavutama survetugevuse vähemalt 17,0 MPa (vt. Joonis 3.3), kuid Eestis kasutatakse vastavalt standardile EVS-EN 197-1:2011 palju suurema algfaasis saavutatava survetugevusega portlandtsemente [8, 19].

## 4. Tulemused

Treenimisjärgsel salvestamisel valiti parimate hüperparameetritega mudel. GridSearch algoritmi abil leiti RFR mudeli jaoks järgmised hüperparameetrid:

```
bootstrap: True                min_samples_split: 2
ccp_alpha: 0.0                min_weight_fraction_leaf: 0.0
criterion: squared_error      monotonic_cst: None
max_depth: None               n_estimators: 150
max_features: 0.5            n_jobs: None
max_leaf_nodes: None         oob_score: False
max_samples: None            random_state: 0
min_impurity_decrease: 0.0   verbose: 0
min_samples_leaf: 1          warm_start: False
```

DTR-mudeli jaoks:

```
ccp_alpha: 0.0                min_samples_leaf: 1
criterion: squared_error      min_samples_split: 5
max_depth: None               min_weight_fraction_leaf: 0.0
max_features: None            monotonic_cst: None
max_leaf_nodes: None         random_state: 0
min_impurity_decrease: 0.0   splitter: best
```

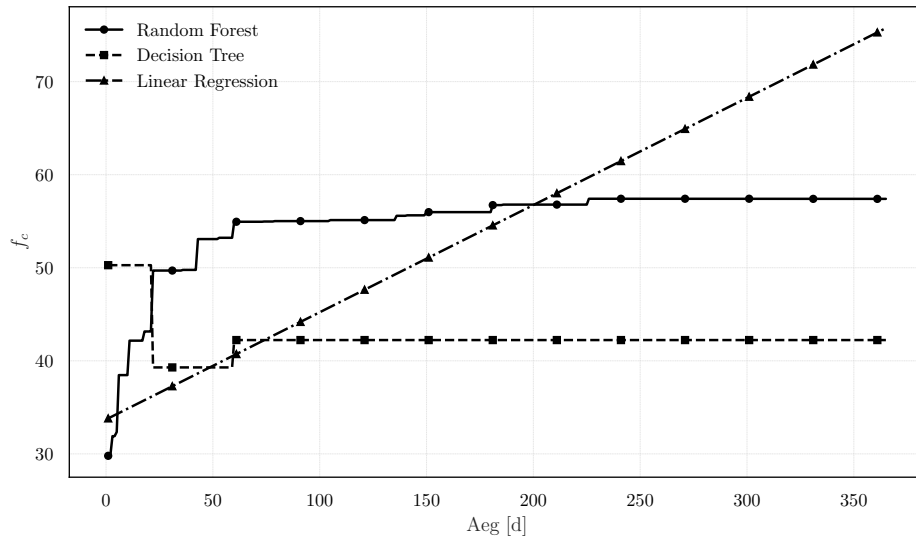
Ning LR-mudeli jaoks:

```
copy_X: True                  positive: False
fit_intercept: True           tol: 1e-06
n_jobs: None
```

Tabel 4.1: Mudelite veahinnangute väärtused

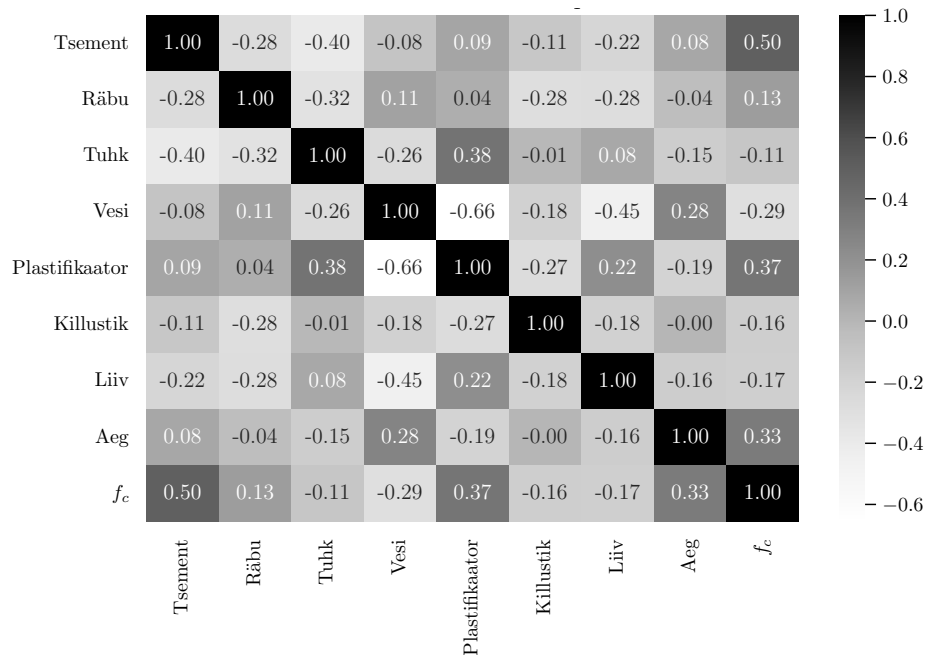
	RFR-mudel	DTR-mudel	LR-mudel
$MSE(\hat{y}_i)$	19,4	53,9	95,6
$R^2$	0,9	0,8	0,6
MAE	3,1	4,6	7,9

Mudelite ennustuste ruutkeskmised vead (ingl. k. *Mean Square Error* (MSE)), keskmised absoluutvead (MAE) ning determinatsioonikordajad  $R^2$ . Determinatsioonikordaja näitab kui suurt osa andmetest suudab mudel kirjeldada.  $R^2 \in [0; 1]$ . RFR-mudeli otsustuspuude keskmine sügavus on 18,2 haru. Mudeli kõikide otsustuspuude sügavus jääb vahemikku 15–22 haru. Leitud veahinnangutest on näha, et kõige väiksemat viga teeb ennustamisel RFR-mudel,



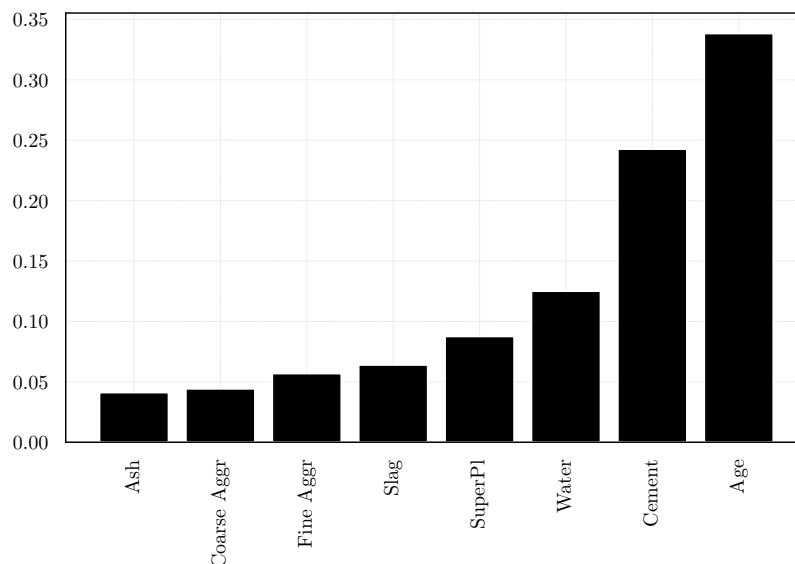
Joonis 4.1: Treenitud mudelite ennustatud survetugevuse võrdlus ühe retsepti alusel sõltuvana ajast. Segu retsepti vt 1. reast Tabelis 3.1. Graafiku väärtused on ennustatud sammuga 1 ööpäev.

millele järgnevad vastavalt vea kasvule DTR-mudel ja LR-mudel (vt Tabel 4.1). Bakalaureuse töö raames treenitud mudelite kasutamiseks vajalikud .pkl-failid on leitavad Lisast A.2.

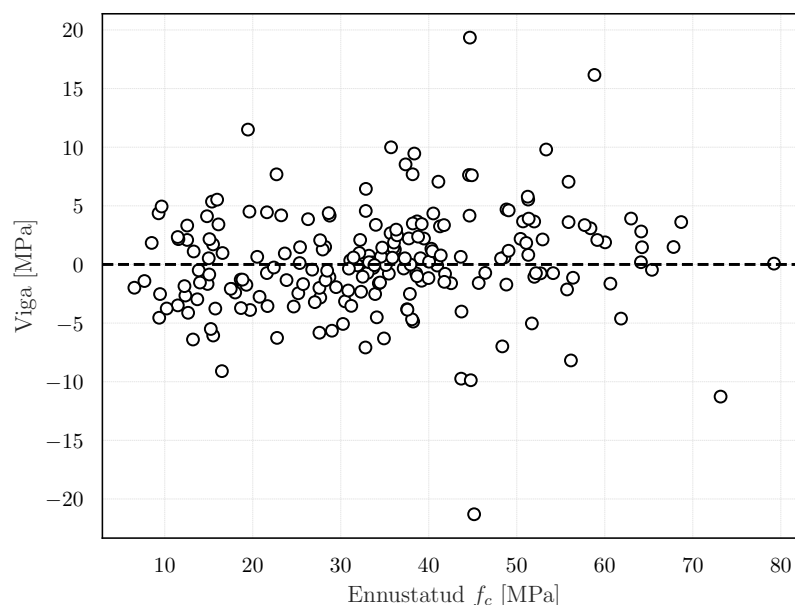


Joonis 4.2: Andmekogu atribuutide korrelatsioonimaatriks, millel mudelit treeniti

Joonisel 4.4 on näha, et enamik RFR-mudeli ennustusi langeb vahemikku  $\pm 5$  MPa, kuid esineb ka teatud määral erandlikke juhtumeid, kus mudeli viga ennustusel ületab koguni 20MPa. Arvestades, et üks selline suur viga on tehtud ennustusega umbes 45 MPa, tähendab see, et tegeliku betoonisegu survetugevus on üle 60 MPa. Sellest järeldub, et mudel on prognoosinud antud betoonisegu survetugevuse rohkem kui 35% väiksemaks kui see tegelikult on.



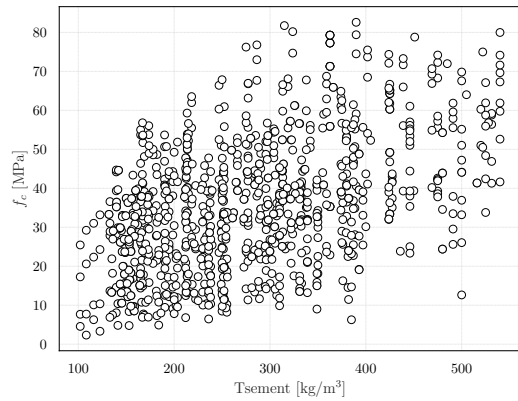
Joonis 4.3: Koostisosade tähtsus mudeli ennustamisel



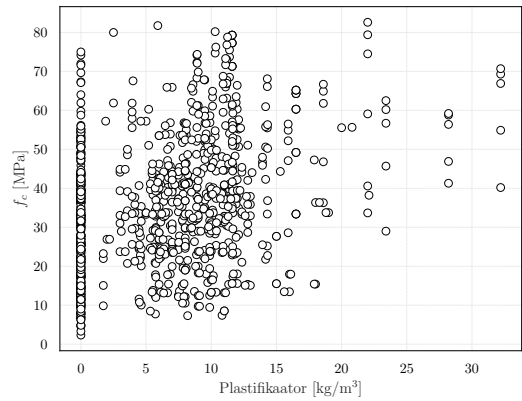
Joonis 4.4: Mudeli jäägi  $e_i = y_i - \hat{y}_i$  diagramm, kus  $y_i$  on kontrollandmete tegelik väärtus ning  $\hat{y}_i$  mudeli punkthinnang prognoositavale väärtusele.

Jooniselt 4.1 on näha, et RFR-mudeli ennustused vihjavad betooni survetugevuse logaritmilisele muutusele ajas. Samuti on näha, et üksik otsustuspuu regressioon annab vea mõttes keskmiselt täpsema tulemuse, kui lineaarne regressioon, kuid selle alusel betoon on kõige tugevam segu valamisel, mille järel betooni survetugevus ajas kahaneb. LR-mudel suudab tabada ära vaid üldise füüsilise seose: ajapikku segu muutub tugevamaks. Sellegi poolest ei suuda mudel luua tulemust, millega on võimalik adekvaatselt hinnata segu survetugevuse kasvu kiirust erinevate perioodide vältel.

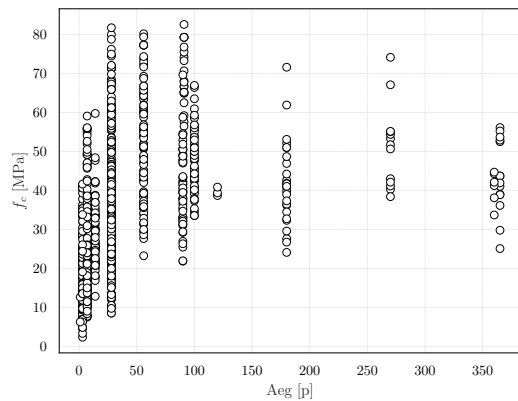
Mudelist poolne betoonisegu survetugevuse alahindamine on turvalisuse poolest parem ennustuse viga, kuid selle negatiivseteks külgedeks on potentsiaalne suurem majanduslik kulu ning



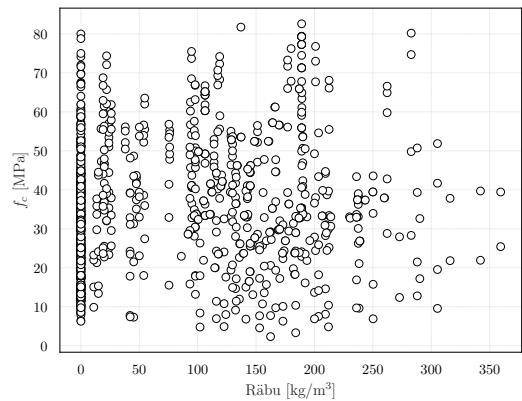
(a) Tsement



(b) Superplastifikaator



(c) Tardumisperiod



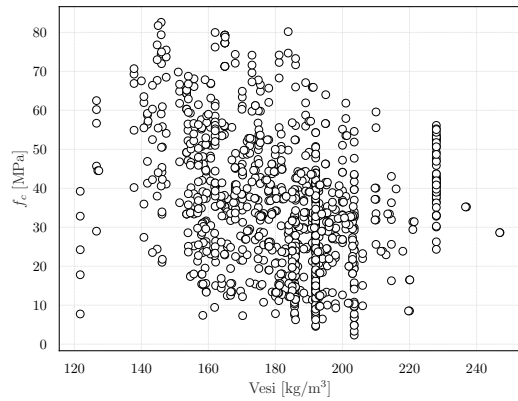
(d) Räbu

Joonis 4.5: Survetugevusega positiivselt korreleeruvad atribuudid

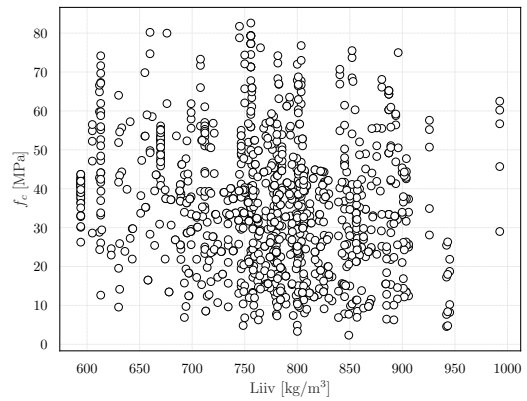
ulatuslikum ölooloogiline jalajälg. Mudelipoolne betoonisegu tugevuse ülehindamine võib aga tuua kaasa konstruktsiooni purunemist, millel on oluliselt halvem tagajärg nii majanduslikust kui ka ohutuse vaatepunktist. Hädavajalik on mudeli mõjukate erindite täpsem analüüs. Ehituses kasutatavad varutegurid on tihti lähedal 1,5-le või suuremale väärtusele olenevalt konstruktsiooni otstarbest [1], mistõttu pruugivad need staatilise koormuse korral erindid ära katta. Selline olukord, siiski, ei ole ideaalne, kuna varuteguri väärtus on mõeldud arvestama sisse ka muid ettearvamatuid konstruktsiooni koormusi.

Andmekogu [18] rakendatud Pearsoni korrelatsiooni analüüsi, mille käigus ilmsid atribuutide vahel seosed, mida kirjeldab Joonisel 4.2 olev korrelatsioonimaatriks. Maatriksist järeldub, et positiivne korrelatsioon betoonisegu survetugevusega on tsemendil (*Cement*, 0,50), superplastifikaatoril (*SuperPl*, 0,37), segu tardumisperiodil (*Age*, 0,33) ja räbu kogusel segus (*Slag*, 0,13). Samuti ilmneb, et negatiivne korrelatsioon survetugevusega on vee kogusel (*Water*, -0,29), peentäitematerjalil (*Fine Aggr*, -0,17), jämetäitematerjalil (*Coarse Aggr*, -0,16) ning tuha kogusel segus (*Ash*, -0,11).

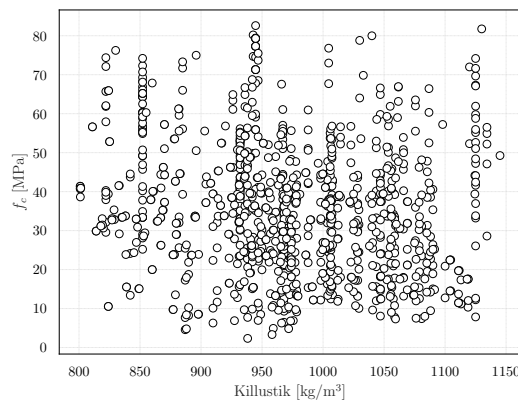
Jooniselt 4.5 on näha, et andmekogus esitatud andmed ei võimalda terviklikult määrata survetugevuse  $f_c$  ajadünaamikat (vt Joonis 4.5c). Analoogselt on Joonisel 4.6 näha, et korrelatsioon on vaid osaline tuha korral ning kindlast vee hulgast alates hakkab korrelatsioon samuti



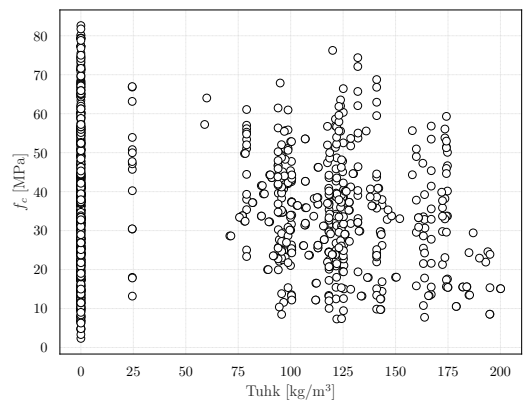
(a) Veski



(b) Liiv



(c) Killustik

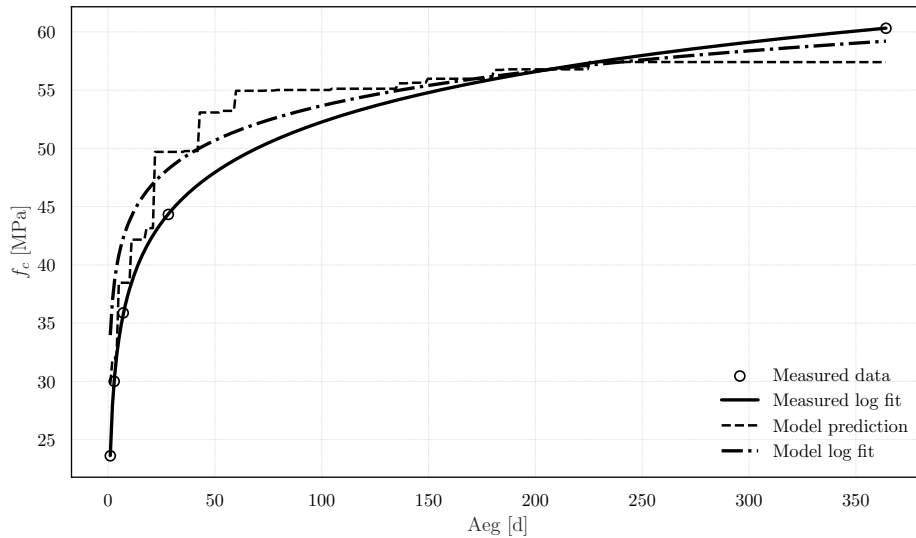


(d) Tuhk

Joonis 4.6: Survetugevusega negatiivselt korreleeruvad atribuudid

nõrgenema (vt vastavalt Joonised 4.6d, 4.6a). Joonisel 4.5c on näha, et aja korrelatsiooni ei ole andmekogu lõikes võimalik selgelt näha, kuid mudeli treeningalgoritm suutis siiski korrelatsiooni tuvastada ning määras selle kõikidest atribuutidest kõige olulisemaks (vt ka Joonis 4.3). Tsemendi korrelatsioon on seevastu andmekogus selgelt eristuv (vt Joonis 4.5a), millest järeldub, et tsemendi kogus segus on väga tugevalt seotud betoonisegu survetugevusega. Järeldust kinnitavad ka peatüki 1 viidete allikad. Vee olulisus segus (Joonis 4.3) on samuti tähtsal kohal, kuid andmekogu lõikes on näha selle palju väiksemat mõju survetugevusele. See aga ei tähenda vee vähest olulisust betoonikeemia protsessides. Kuna betoon vajab pidevat niiskust tardumisperioodi jooksul, ei ole võimalik tekitada andmekogus vee koguse suurt varieeruvust, mistõttu analüüs pruugib vee olulisust liigselt alahinnata. Sellegi poolest joonisel 4.3 on näha, et tardumisperiood, tsemendi kogus ja vee kogus suudavad kirjeldada ära ligi 70% andmekogust.

Räbu (ingl. k. *Slag*, joonis 4.5d) korrelatsioon on positiivne, kuid nõrk. Räbu on pozzolaanne materjal, mis soodustab C-S-H teket segus. Korrelatsioonist järeldub, et räbu õige doseerimine võib tõsta betooni survetugevust, kuid selle olemasolu segus ei mängi primaarset rolli survetugevuse kasvuks. Vastavat ajaolu kinnitavad ka mitmed allikad. Tuha korrelatsioon survetugevusega on samuti nõrk (Joonis 4.6d). [2, 6]



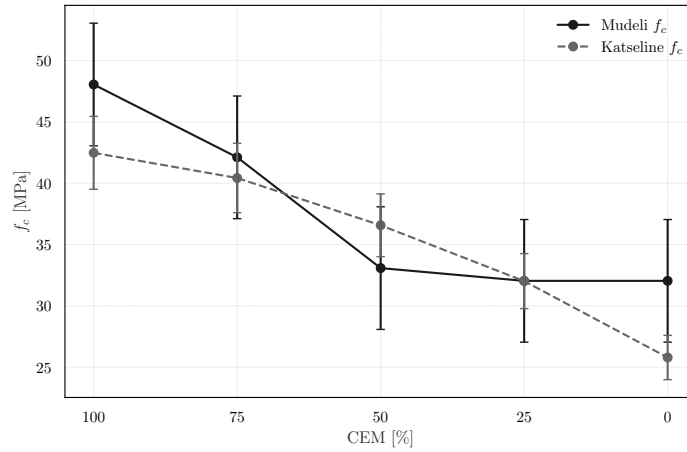
Joonis 4.7: Portlandtsemendi segu tsemendi-vee suhtega 0,5. Mõõdetud tulemus ja täpsem retsept laenatud allikast [23]

Negatiivselt korreleeruvatest atribuutidest selgelt paistab välja vee korrelatsioon survetugevusega (vt Joonis 4.6a). Liiga vähe või liiga palju vett põhjustab betooni survetugevuse langust. Vastav korrelatsioon on kooskõlas ka varasemate uuringute ning õpikute teoreetiliste alustega. Liiga vähese vee korral ei teki hüdratsiooniprotsessis piisavalt C-S-H geeli ning toimub kuivsettimine, mistõttu on betoon lõpuks rabedam. Liigne vesi põhjustab aurustumisel betoonisegu liigse poorsuse ning tekitab ettearvamatuid pingekontsentraatoreid segu sisse, mis viivad betooni madalama purunemiskiirini. Joonisel 4.7 on näha mudeli ennustust sõltuvana ajast ühe kindla segu koostise järgi. [2, 6, 24]

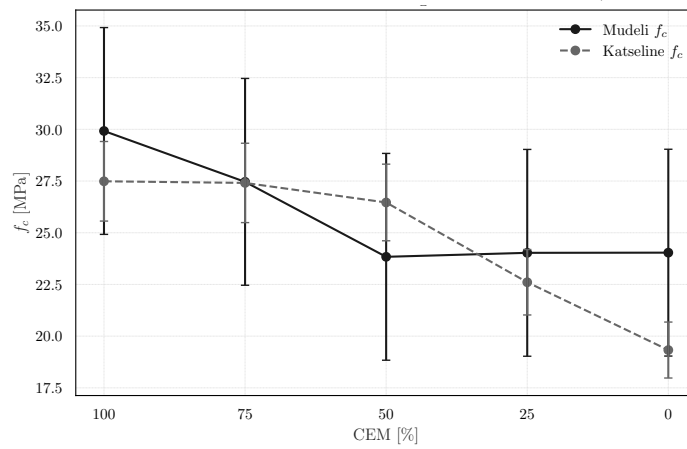
Tuha kasutamisest betoonisegus on tehtud täiendavaid uuringuid. Tallinna Tehnikaülikoolis kaitstava magistr töö [4] raames on tehtud uuring tsemendi ja kohalike elektrijaamade põlevkivituhha kasutamisest betoonisegu valmistamisel. Joonisel 4.8 on välja toodud 3 erinevat retsepti, milles kasutati erinevat plastifikaatori kogust ning tuha ja tsemendi osakaalu. Samadel joonistel (4.8b, 4.8c, 4.8a) on välja toodud ka mudelipoolsed survetugevuse ennustused täpsusega 5 MPa ( $RMSE(\hat{y}_i) \approx 4,4$  MPa (ümardatud turvalisuse kaalutlustel täisarvuni)). Tehtud segude survetugevuste täpsem analüüs ning segude valmistamise meetoodika on kirjeldatud vastavas magistr töö [4]. Segude erinevused, mis on olulised käesolevas töös, on toodud mainitud vastavate graafikute juures Joonisel 4.8.

Joonistel 4.8a ja 4.8b on näha, et mudel suudab ennustada konstantsete plastifikaatorite kogustega segude survetugevust peaaegu täielikult mudeli saadud  $RMSE(\hat{y}_i)$  piires. Joonisel 4.8a 25% tsemendi sisalduse korral (põlevkivi tuhka on vastavalt 75%) langevad mõõtevea piires täielikult kokku. Sama olukord esineb ka joonisel 4.8b 75% tsemendi sisalduse korral segus. Joonisel 4.8b jäävad kõik survetugevuse tegelikud väärtused RFR-mudeli  $RMSE(\hat{y}_i)$  vahemikku.

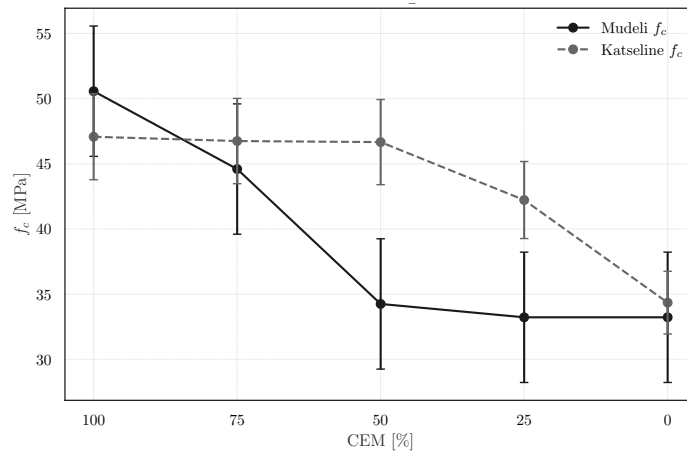
Joonisel 4.8c esineb aga 50- ja 25-protsendilise tsemendi sisaldusega segude korral survetugevuste erinevus ligi 25%. Selline survetugevuste erinevus pruugib olla tingitud mudeli arusaamast



(a) Muutuva tsemendi ja põlevkivi tuha osakaaluga segu ning konstantse superplastifikaatori kogusega 3,49979 kg/m<sup>3</sup>



(b) Muutuva tsemendi ja põlevkivi tuha osakaaluga segu ning konstantse superplastifikaatori kogusega 4,64972 kg/m<sup>3</sup>



(c) Muutuva superplastifikaatori, tsemendi ja põlevkivituhha sisaldusega segu. CEM% langusele vastab superplastifikaatori kogus segus järgmiselt: [0; 0; 0,32592; 0,76466; 1,08817] kg/m<sup>3</sup>

Joonis 4.8: Magistritöö raames tehtud betoonisegude survetugevuse mõõdetud ja mudeli poolt ennustatud väärtused

vähese plastifikaatori koguse ja suure tuhasisalduse korrelatsioonist. Joonise 4.8c segu korral on suurim plastifikaatori kogus ligi 1,1 kg, mis on 4 korda vähem kui segus 4.8b ning ligi 3 korda

vähem, kui segus 4.8a. Treeningandmekogu alusel pruukis mudel seostada vähest plastifikaatori kogust ning suurt tuhasisaldust vähese survetugevusega (vt korrelatsioone jooniselt 4.2 ning atribuutide jaotusi andmekogus joonistelt 4.5a ja 4.6d). Lisaks on tuleb superplastifikaatori koguse korral kasutada rohkem vett. Vee koguse korrelatsioon survetugevusega on samuti mudeli andmekogu lõikes negatiivne (vt Joonis 4.6a).

## 5. Arutelu

Nagu selgub ka kirjandusest, on betooni protsessi mittelineaarsus oluliselt keerukam ning soovituslik on kasutada ennustamiseks ANN või DNN masinõppe mudeleid. Arvestada tuleb ka asjaoluga, et täpsema mudeli treening nõuab kvaliteetsemaid treeningandmeid. St treeningandmeid, milles on rohkem survetugevust mõjutavaid parameetreid korrektsel kujul ning rohkem andmeridu, sh andmeridu, mis sisaldavad ekstreemumeid ja erijuhte. Treenitud mudeli kasutus võib tuua kasu ettevõtetele, mis tegelevad igapäevaselt betooni valamisega. Igapäevase betoonisegude valmistamise käigus on ettevõtetel võimalik lisada oma andmeid ettevõttesisestesse andmekogudesse ning kasutada neid mudeli täpsemaks treenimiseks.

## Kokkuvõte

Käesoleva bakalaureusetöö eesmärgiks oli uurida Pythoni programmeerimiskeeles kirjutatud masinõppe mudelite suutlikkust tuvastada andmestikest betooni survetugevust mõjutavate tegurite omavahelisi korrelatsioone ning suutlikkust nende põhjal ennustada betoonisegu survetugevust. Ennustus toimus etteantud koostise järgi, millesse kuulub 8 parameetrit: tsemendi kogus, vee kogus, räbu kogus, tuha kogus, superplastifikaatori kogus, peentäitematerjali ehk liiva kogus, jämetäitematerjali ehk killustiku kogus ning tardumisperioodi pikkus päevades. Mudelite treeninguks kasutati I. C. Yeh' poolt loodud tuntud masinõppeks adapteeritud andmestikku. Töös käsitleti LR-mudelit, DTR-mudelit ning RFR-mudelit. Analüüsitud mudelitest kõige väiksema veaga ennustusi tegi RFR- mudel, mille ruutkeskmine viga oli ligi 19,4 MPa. Sellele järgnes matemaatilise täpsuse poolest üksiku otsustuspuu mudel, mille ruutkeskmine viga oli 53,9 MPa. Paraku oli mudeli viga puhtmatemaatiline ning üksiku otsustuspuu mudel ei suutnud adekvaatselt tuvastada sisendparameetrite koosmõju survetugevusele, mistõttu andis see mudel füüsikaliselt vale tulemus. Kõige ebatäpsemini ennustas lineaarse regressiooni mudel, mille ruutkeskmine viga oli 95,6 MPa.

Juhusliku metsa mudel, olles kõige täpsem analüüsitud mudelitest, siiski andis tegelikust väärtusest ligi 25% võrra vale tulemus segude puhul, kus oli suur tuha ning väike plastifikaatori osakaal segus. Mudeli hindamisel teostati Pearsoni korrelatsiooni analüüs ning vaadati mudeli hinnanguid parameetrite olulisuse osas, mille tulemusena suutis mudel omistada parameetritele suhteliselt adekvaatsed olulisuse näitajad.

Mudeli kontrollimisel kasutatud vastava magistr töö raames saadud andmed näitasid, et mudel on laias pildis usaldusväärne ning suudab ennustada betoonisegude survetugevusi täpsusega ligi 10%, kuid esineb üksikuid suuremaid kõrvalekaldeid. Väga suure ebatäpsusega tulemuste tõttu on vaja viia läbi erijuhtude sügavam analüüs, et leida ennustuste vigade põhjuseid.

## Tänuavaldused

Käesoleva töö autor soovib tänada antud töö ilmumisse panustanud isikuid ja institutsioone. Autor soovib tänada oma põhijuhendajat, teadur Dmitri Kartofelevit, kes juhendas töö autorit iga nädal: aitas leida koodis vigu, tõlgendada lahti töö tulemusi ning oli toeks autorile igal etapil. Autor soovib tänada oma kaasjuhendajat, vanemteadur Heiko J. Herrmanni, tänu kellele tuli teemaks taolise töö läbiviimine, tekkis koostöö Tallinna Tehnikaülikooli Ehituse ja Arhitektuuri Instituudi kolleegidega ning kes aitas leida hulgaliselt teaduslikke materjale tõstmaks autori pädevust betooni valmistamise teoreetilistes alustõdedes. Autor soovib tänada Ehituse ja Arhitektuuri Instituudist koostööd teinud lektor Tanel Tuisku. Tema suunavad küsimused aitasid autoril mõttestada paremini lahti töö tulemusi, tema sügavad teadmised antud valdkonnas aitasid autoril kontrollida mudeli töökindlust praktiliste katsete kaudu ning tänu temale sündis ka koostöö magistritööd kaitsva tudengiga. Autor soovib tänada doktorant-nooremteadurit Mattias Põldarut, kelle suunavad küsimused aitasid autoril omandada sügavamad teadmised betoonisegu koostisosade ning võrreldavuse kohta. Viimaks soovib autor tänada magistritöö autorit, Rasmus Rahuojat, kes lubas kasutada käesoleva töö raames mudeli kontrollimiseks tema magistritöös esitatud andmeid.

## Kasutatud kirjandus

- [1] Akrediteerimiskeskus, E. S. ja. *EVS-EN 1992-1-1:2005 Eurokoodeks 2: Betoonkonstruktsioonide projekteerimine. Osa 1-1: Üldreeglid ja reeglid hoonetele*. Tehniline raport. Tallinn, Estonia: Eesti Standardimis- ja Akrediteerimiskeskus, 2005.
- [2] Marchewka, A., Ziolkowski, P. ja Galán, S. G. „Modeling of strength of high-performance concrete using artificial neural networks“. *Cement and Concrete Research* 15 (1 1998). DOI: 10.1038/S41598-025-97060-W.
- [3] Shaaban, M., Amin, M., Selim, S. ja Riad, I. M. „Modeling of strength of high-performance concrete using artificial neural networks“. *Cement and Concrete Research* 15 (1 1998). DOI: 10.1038/S41598-025-10342-1.
- [4] Rahuoja, R. „Põlevkivituhast valmistatud betooni mahumuutused. Betooni töödeldavuse muutus ajas sõltuvalt sideainest“. Magistritöö. Tallinna Tehnikaülikool, 2026.
- [5] Yeh, I. C. „Modeling of strength of high-performance concrete using artificial neural networks“. *Cement and Concrete Research* 28 (12 1998), lk. 1797–1808. DOI: 10.1016/S0008-8846(98)00165-3.
- [6] Neville, A. M. *Properties of Concrete*. 5. väljaanne. Pearson Education, 2011.
- [7] European Committee for Standardization. *EN 197-1: Cement – Part 1: Composition, specifications and conformity criteria for common cements*. Tehniline raport. Brussels, Belgium: European Committee for Standardization, 2011.
- [8] Eesti Standardimis- ja Akrediteerimiskeskus. *EVS-EN 197-1:2011 Tsement. Osa 1: Harilike tsementide koostis, spetsifikatsioonid ja vastavuskriteeriumid*. Tehniline raport EVS-EN 197-1:2011. Eesti Standardimis- ja Akrediteerimiskeskus, 2011.
- [9] Wang, H., Wang, Q., Gao, T., Wang, J., Sun, C., Ji, Z., Meng, J. ja Ta, B. „Analysis of strength failure in gangue-based cementitious backfill body from a microscopic perspective“. *PLOS ONE* 19 (2024), e0300102. DOI: 10.1371/journal.pone.0300102.
- [10] American Concrete Institute. *Report on Chemical Admixtures for Concrete*. Tehniline raport. Farmington Hills, Michigan, USA: American Concrete Institute, 2016.
- [11] Wang, W. ja Yue, Q. „The Time Variation Law of Concrete Compressive Strength: A Review“. *Applied Sciences* 2023, Vol. 13, Page 4947 13 (8 2023), lk. 4947. DOI: 10.3390/APP13084947.
- [12] Palanisamy, A. K., Jegatheeswaran, D., Ananth, C., Bhogayata, A., Juneja, B., Buribay, R. ja Mezgebe, T. T. „Prediction of early age compressive strength of concrete using machine learning“. *Scientific Reports* 2025 15:1 15 (1 2025), lk. 45293–. DOI: 10.1038/s41598-025-29233-6.

- [13] Scikit-learn arendajad. *RandomForestRegressor* — *Scikit-learn Documentation*. 2025. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html> (vaadatud 28.04.2026).
- [14] Molnar, C. *Inductive biases of the Random Forest and their consequences*. 2024. URL: <https://mindfulmodeler.substack.com/p/inductive-biases-of-the-random-forest>.
- [15] Witten, I. H., Frank, E. ja Hall, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3. väljaanne. Elsevier, 2011.
- [16] Pihlak, M. *Klassikaline ja mitteparameetriline matemaatiline statistika*. Õpik kõrgkoolidele. Tallinna Tehnikaülikooli Kirjastus, 2016, lk. 78–82.
- [17] Scikit-learn arendajad. *1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking* — *scikit-learn 1.8.0 documentation*. 2026. URL: <https://scikit-learn.org/stable/modules/ensemble.html#adaboost>.
- [18] Yeh, I.-C. *Concrete Compressive Strength* — *UCI Machine Learning Repository*. URL: <https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength> (vaadatud 24.05.2026).
- [19] ASTM. *ASTM C150: Standard specification for Portland Cement*. Tehniline raport. ASTM, 2007.
- [20] Alabi, A. „The Effect of Fineness Modulus of Fine Aggregate (sand) on Concrete Compressive Strength“. Teoses: *International Conference on Science, Technology and Modern Society*. Köide 1. 1. 2023, [1–12].
- [21] Scikit-learn arendajad. *sklearn.model\_selection.GridSearchCV*. 2026. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html) (vaadatud 24.05.2026).
- [22] Eesti Standardimis- ja Akrediteerimiskeskus. *EVS-EN 1008:2002 Mixing water for concrete – Specification for sampling, testing and assessing the suitability of water, including water recovered from processes in the concrete industry, as mixing water for concrete*. Tehniline raport EVS-EN 1008:2002. Eesti Standardimis- ja Akrediteerimiskeskus, 2002.
- [23] Schulze, J. „Influence of water-cement ratio and cement content on the properties of polymer-modified mortars“. *Cement and Concrete Research* 29.6 (1999), lk. 909–915. DOI: [https://doi.org/10.1016/S0008-8846\(99\)00060-5](https://doi.org/10.1016/S0008-8846(99)00060-5).
- [24] Abd Elaty, M. a. a. „Compressive strength prediction of Portland cement concrete with age using a new model“. *HBRC Journal* 10.2 (2014), lk. 145–155. DOI: <https://doi.org/10.1016/j.hbrcj.2013.09.005>.

## Annotatsioon

Käesoleva bakalaureusetöö raames treenitakse Pythoni Scikit-learn teeki kasutades erineviad masinõppe mudeleid ennustama betooni survetugevust vastavalt koostisele ja tardumisperioodi pikkusele piirveaga 10%. Töö jooksul tutvutakse betoonikeemiaga ning koostisosade mõjuga segu survetugevusele. Samuti tutvutakse täpsemini Scikit-learn teegi olemasolevate meetoditega, millega on võimalik teha mudeli treenimine efektiivsemaks saavutades seeläbi kõrgema täpsusega, parem statistiline mudel.

Mudeleid treenitakse kasutades I. C. Yeh' loodud andmekogu, 5-kordset ristkontrolli ja GridSearch algoritmi hüperparameetrite võrke. Seejärel valitakse parimate meetrikutega mudel, mille täpsust kontrollitakse katseandmetega koostöös Tallinna Tehnikaülikooli Ehituse ja Arhitektuuri Instituudiga. Oluline on mitte ainult mudeli statistiline täpsus vaid ka selle suutlikkus saavutada vajalik täpsus kirjeldades korrektselt protsessi sõltuvust ajast.

Kõikidest mudelist parima tulemuse andis juhusliku metsa regressioonimudel suutes otsustada erinevate segude survetugevust sõltuvana ajast keskmise veaga ligi  $\pm 5$  MPa. Üksiku otsustuspuu mudel andis statistilised sarnase tulemuse, kuid mudeli järgi betooni survetugevus ajas kahaneb ning on suurim esimestel päevadel. Lineaarse regressiooni mudel prognoosis tulemusi, mis andis tulemuse täpsusega  $\pm 10$  MPa 1-aastase perioodi jooksul, kuid mudeli alusel betooni survetugevuse kasv jätkus ajas lõpmatuseni. Andmestiku lõikes vaadatakse parameetrite korrelatsiooni Pearsoni korrelatsiooni analüüsi abil. Töö lõpus esitatakse mudeli tööd kirjeldavad graafikud ning autoripoolne hinnang mudelile ja soovitus edaspidisteks uuringuteks.

# Abstract

The aim of the following bachelor's thesis is to train different machine learning models to predict the compressive strength of concrete based on the given recipe and hydration period with a deviation from the real value of no more than 10%. This thesis examines the fundamentals of concrete chemistry and the influence of individual components in regards to concrete compressive strength. An overview of the available machine learning methods included in the Python's Scikit-learn library, through which model training and evaluation can be made more efficient achieving more precise statistical models, is also provided.

The models are trained using the dataset from the University of California, Irvine's machine learning repository, that was contributed by I. C. Yeh, 5-fold cross-validation and GridSearch hyperparameter networks. Models with the best metrics are then selected, accuracies of which are evaluated in collaboration with the Institute of Civil Engineering and Architecture of Tallinn University of Technology. Priority is set not only on models' statistical accuracy but also on models's ability to describe the compressive strength development in relation to curing period.

Random forest regression model's performance was deemed the best among other analysed models, because it could predict the compressive strength with a mean error of  $\pm 5$  MPa. Even though decision tree regression model gave statistically a similar outcome, it showed that concrete is the strongest at it's early stages and loses it's strength over time. Linear regression model predicted strength with an accuracy of  $\pm 10$  MPa throughout a one-year period, however, according to the model the concrete strength growth wouldn't stop over time and continue to develop towards infinity. Pearson's Correlation Analysis is used to determine the correlation between different components used in concrete mixing process, curing period and compressive strength. The end of the thesis contains graphs describing the models' accuracy and author's assessment of the model as well as author's suggestions regarding possible further research.

## A. Lisad

### A.1. Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Mark Dovydovich,

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Betooni survetugevuse prognoosimine kasutades andmekaevet”, mille juhendaja on Dmitri Kartofelev ja kaasjuhendaja on Heiko J. Herrmann
  - 1.1 reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2 üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

25.05.2026

## A.2. Link treenitud mudelite failidele.

Allpool on võimalik alla laadida treenitud LR- DTR- ning RFR-mudelite kasutamiseks vajalikud binaarsed .pk1 failid.

<https://www.dropbox.com/scl/fo/5z38x48tv38g8f7rd17tk/ADir6VqYtS4Jg16f4E9w4yk?rlkey=3gt2fbvzsyzgucgofv6vxsz44&st=awnome5q&dl=0>