



Keelemudelite ja keele tehnoloogia vahendite kasutamine eesti kui emakeele kirjandite hindamisel

Merilin Aruvee, Kais Allkivi, **Andres Karjus**,
Krister Kruusmaa, Katarin Leppik, **Silvia Maine**,
Taavi Kamarik, Harli Kodasma

Peaküsimused

- 1) Kui võimekad on statistilise keeletöötuse vahendid ja generatiivsed keelemudelid eesti keele kirjandite hindamisel?
- 2) Kuivõrd erinevad inimhindajate ja e-tööriistade hinded?



Taust ja kontekst

- 2027 põhikooli e-eksam, st e-kirjand
- Katseeksamid toimunud 2 a
- Kirjutatakse alusteksti põhjal
- Hindamine toimub keskkonnas EIS, piiratud võimetega hindamistööriist

Kirjuta kirjand vastuse aknasse. Pealkiri vorminda paksus kirjas ja joonda keskele. Reavahetus alustab uut lõiku, seepärast ei ole tekstis vaja kasutada taandridu.

→ Pealkirjasta kirjand.
→ Kirjuta sissejuhatus, kus püstitad probleemi ja sõnastad enda põhiseisukoha.
→ Toeta oma põhiseisukohta vähemalt kahe teemaarendusiõiguga.
→ Igas lõigus esita alaväide, milles kasutad vähemalt ühte alusteksti näidet. Lisa ka enda näiteid.
→ Kokkuvõttes esita peamised järeldused.
→ Ülesande täitmisel võid kasutada [Sõnaveebi](#) portaali. Lingil klõpsates avaneb see uues aknas.

B I [font size icons] [bullet point icon] [list icon] 0 sõna

Kirjuta kirjand siia aknasse.

Uurimismaterjal (2024. a katseeksam)

- 795 põhikooli kirjandit (9. klass)
 - Statistiline keeletöötlus, keelemudelid
- 764 gümnaasiumi kirjandit (12. klass)
 - Statistiline keeletöötlus
- Kahe eksperdi antud hinded, kellest üks Harno ekspert, teine õpilaste õpetaja



Hindamisaspektid, iga kriteerium annab 0-3 punkti

Sisu ja ülesehitus, 12 punkti

Pealkiri ja sissejuhatus

Teemaarendus: alusteksti kasutamine

Teemaarendus: argumentatsioon

Lõpetus

Lausestus ja sõnavalik, 6 punkti

Sõnavalik

Lausemoodustus (ühildumine, sõnajärg, reksioon)

Õigekiri ja vormistus, 9 punkti

Kirjavahemärgistus

Õigekiri ja grammatika

Liigendus ja vormistus



SISU JA ÜLESEHITUS (maksimaalne tulemus 12 punkti)

Pealkiri ja sissejuhatus		Teemaarendus <u>Alusteksti kasutamine</u>		Teemaarendus <u>Argumentatsioon</u>		Lõpetus	
3	✓ Pealkiri on haarav või kitsendab teemat ja seostub selgelt probleemipüstitusega ✓ Sissejuhatuses avatakse probleemi taust ✓ Probleemipüstitusena esitatakse üks selgelt sõnastatud põhiväide või -küsimus, mis loob aluse teemaarenduseks	3	✓ Alusteksti(de)st on toodud vähemalt üks näide ✓ Alustekstile on osundatud tsitaadi või refereeringuga ✓ Alusteksti näidet laiendatakse enda mõtetega	3	✓ <u>Alaväited</u> on selgelt sõnastatud ja seotud <u>põhiväitega</u> ✓ <u>Alaväited</u> on esitatud loogilises järjekorras ✓ Igas lõigus on esitatud näiteid selgitatud ja laiendatud, need on asjakohased ✓ Näidetest kasvavad loogiliselt välja lõikude järeldused	3	✓ Sissejuhatuses püstitatud probleemile on vastatud ✓ Lõikude peamised järeldused on teises sõnastuses kokku võetud
2	✓ Pealkiri seostub probleemipüstitusega ✓ Sissejuhatuses avatakse probleemi taust ✓ Probleemipüstituses esitatud põhiväide või -küsimus haakub teemaarendusega osaliselt või avatakse see üldsõnaliselt	2	✓ Alusteksti(de)st on toodud vähemalt üks näide ✓ Alustekstile on osundatud ebaselgelt (nt tekstile või autorile ei viidata korrektselt) ✓ Alusteksti näidet laiendatakse enda mõtetega	2	✓ <u>Alaväited</u> on selgelt sõnastatud ja osaliselt seotud <u>põhiväitega</u> ✓ <u>Alaväited</u> on esitatud loogilises järjekorras ✓ Igas lõigus on esitatud näiteid selgitatud ja laiendatud, need on asjakohased ✓ Näidetest ei kasva selgelt välja järeldused	2	✓ Sissejuhatuses püstitatud probleemile on vastatud ✓ Lõikude peamisi järeldusi korratakse sissejuhatusesele ja teemaarendusele ligilähedases sõnastuses
1	✓ Pealkiri on üldsõnaline või puudub ✓ Probleemi taust avatakse osaliselt ✓ Probleemipüstituses esitatud põhiväide või -küsimus on ebaselgelt sõnastatud, endastmõistetav või tõestamatu	1	✓ Alusteksti(de)st on toodud vähemalt üks näide ✓ Alustekstile on osundatud temaatiliselt, ilma viitamata ✓ Alusteksti näidet ei seota enda mõtetega piisavalt selgelt	1	✓ Kõik <u>alaväited</u> pole selgelt sõnastatud ✓ <u>Alaväited</u> on osaliselt seotud <u>põhiväitega</u> ✓ Toodud näiteid on osaliselt laiendatud ✓ Näidetest ei kasva välja järeldused	1	✓ Sissejuhatuses püstitatud probleemile on vastatud osaliselt või tuuakse sisse uus teema ✓ Lõikude peamisi järeldusi korratakse sissejuhatusese ja teemaarendusega samas sõnastuses

LAUSESTUS JA SÕNAVALIK (maksimaalne tulemus 6 punkti)

Sõnavalik		Lausemoodustus (ühildumine, sõnajärg, rektsoon)	
3	✓ Sõnavalik on isikupärane ja rikkalik ✓ Sõnavalik sobib kirjakeelsesse teksti ✓ Esineb üksik sõnastusraskus või -viga	3	✓ Laused on arusaadavad ja terviklikud ✓ Kasutatakse sidusaid ja erineva ülesehitusega lauseid ✓ Esineb üksik lausestuseksimus
2	✓ Sõnavalik on mitmekesine, esineb üksikuid sõnakordusi ✓ Sõnavalik sobib kirjakeelsesse teksti ✓ Esineb mõnesid sõnastusraskusi	2	✓ Laused on arusaadavad ja terviklikud ✓ Kasutatakse sidusaid ja sarnase ülesehitusega lauseid ✓ Esineb mõnesid lausestuseksimusi
1	✓ Sõnavalik on ühekülgne, esineb palju sõnakordusi ✓ Sõnavalik sobib suuremalt jaolt kirjakeelsesse teksti ✓ Esineb palju sõnastusraskusi	1	✓ Laused on suuremalt jaolt arusaadavad ✓ Kasutatakse suuremalt jaolt sidusaid lauseid, mille ülesehitus on ühekülgne ✓ Esineb palju lausestuseksimusi
0	Sõnavalik ei sobi kirjakeelsesse teksti ja sõnastusraskuste tõttu on tekst arusaamatu	0	Laused on ebaselged ja välja arendamata, lausestuseksimuste tõttu on tekst arusaamatu

ÕIGEKIRI JA VORMISTUS (maksimaalne tulemus 9 punkti)

Kirjavahemärgistus (koma, jutumärgid, kriipsud, koolon, lauselõpumärgid)		Õigekiri ja vormimoodustus (algustäht, kokku- ja lahkukirjutamine, häälikuortograafia, käänamine-pööramine)		Liigendus ja vormistus	
3	✓ Kirjavahemärgistus on täpne ✓ Võib esineda 0-2 viga kokku	3	✓ Õigekiri ja vormimoodustus on korrektne ✓ Võib esineda 0-2 viga kokku	3	✓ Tekst on liigendatud ✓ Lõigud on proportsionaalsed ✓ Tekst on trükitud korrektselt, võib esineda 0-2 trükiviga
2	✓ Kirjavahemärgistus on valdavalt täpne ✓ Võib esineda 3-4 viga kokku	2	✓ Õigekiri ja vormimoodustus on valdavalt korrektne ✓ Võib esineda 3-4 viga kokku	2	✓ Tekst on liigendatud ✓ Esineb üksik ebaproportsionaalne lõik ✓ Tekst on trükitud valdavalt korrektselt, võib esineda mõningaid trükivigu (3-4)

Statistiline keeletöötlus

- Juhitud masinõpe: regressioonmudelid eri alamskooride jaoks
- Keelekasutust iseloomustavad kvantitatiivsed tunnused
 - Grammatika, sõnavara, teksti keerukus, keelevead
- Seosed teksti tunnuste ja hindajate skooride vahel → ennustused
- Klassikaline lähenemine teksti automaathindamises: läbipaistev ja väikse arvutuskuluga, sobib osade aspektide hindamiseks

Statistiline keeletöötlus: tarkvara

- Stanza keeletöötlushakett: teksti sõnestamine, lausestamine, lemmatiseerimine, sõnaliigi- ja vormituvastus (Qi et al., 2020)
- Kiirlugemise tarkvara: abstraktsete ja harvaesinevate sõnade tuvastus (Siiber, 2018; 2025)
- Llama-põhine grammatikakontroll (Luhtaru et al., 2024)
- Sarnaseid keeletunnuseid on kasutatud E2 keeleoskustaseme (A2–C1) määramiseks ELLE Tekstihindajas (Allkivi et al., 2024)

Perspektiiv: ELLE Tekstihindaja kohandamine E1 kirjutiste hindamiseks

KORREKTUUR

TASEMEHINNANG

KEERUKUS

SÕNAVARA

ÕIGEKIRI

GRAMMATIKA

Mina hirmsasti armastan lugeda ajakirju. Mul kodus **need** on väga palju. Tihti ma loen ajakirju, mis **on** ajaloost või bioloogiast. Need on väga huvitavad. Kõige huvitavam teema, mida ma olen lugenud **oli** **piratidest**. Kuidas nad oma laevad tegid ja kuidas nad **elasin**. See oli väga huvitav lugeda.

ANALÜÜSI



Tasemete värvikoodid:

-A2

-B1

-B2

-C1

Üldhinnang

A2 - 100%



Teksti keerukus



Grammatika



Sõnavara



TALLINNA ÜLIKOOL
Digitehnoloogia
instituut



TALLINNA ÜLIKOOL
Humanitaarteaduste
instituut

Statistiline keeletöötlus: töövoog

- Seni E2 tasemehinnanguks kasutatud tunnuste **sisuline seotus** e- kirjandite hindamisaspektidega >
- **Korrelatsioonid** eksperthindajate antud skooridega >
- **Masinõppe algoritmide katsetamine** sobivate tunnustega >
- **10-kordne ristvalideerimine:** masina ennustuste keskmine erinevus inimeste hinnangust



Tunnuste seosed kirjandis hinnatavate aspektidega (1)

Hindamis-kriteerium	Keskmine/tugev seos (>0.3) skooriga nii 9. kui 12. klassi kirjandites	Erinevused
Sõnavalik	Sõnavara mitmekesisuse indeksid MTLTD ja RTTR (juuritud lemmade-sõnade suhtarv)	Kobarparanduste arv (9. kl), sõnaasenduste arv (9. kl), keskmine sõnapikkus tähtedes (9. kl), lemmade arv tekstis (12. kl), sõnavara mitmekesisuse indeksid Maas (9. kl) ja CVV (korrigeeritud verbi-varieeruvus; 12. kl)
Lausemoodustus	Kobarparanduste arv	Keskmine sõnapikkus tähtedes (9. kl), harvaesinev sõnavara (9. kl)



Tunnuste seosed kirjandis hinnatavate aspektidega (2)

Hindamisaspekt	Keskmine/tugev seos (>0.3) skooriga nii 9. kui 12. klassi kirjandites	Erinevused
Kirjavahemärgistus	Puuduva kirjavahemärgi paranduste arv ja osakaal	Liigse kirjavahemärgi parandused (12. kl)
Õigekiri ja vormimoodustus	Sõnaasenduste arv, kokku- lahkukirjutusvigade arv, kobarparanduste arv, paranduste arv kokku, paranduste arv sõna ja lause kohta	Õigekirjavigade arv ja õigekirjaveaga sõnade osakaal (9. kl), kobarparanduste osakaal (9. kl)



Keskmine erinevus hindajate skoorist: 9. klass

Hindamise aspekt	M – H	H1 – H2
ÕIGEKIRI JA VORMISTUS (0–9 p)	0,80	1,20
Kirjavahemärgistus (0–3 p)	0,42	0,55
Õigekiri ja vormimoodustus (0–3 p)	0,505	0,60
Liigendus ja vormistus (0–3 p)	0,44	0,58
LAUSESTUS JA SÕNAVALIK (0–6 p)	0,75	1,14
Sõnavalik (0–3 p)	0,41	0,62
Lausemoodustus (ühildumine, sõnajärg, rektsioon) (0–3 p)	0,44	0,64



Keskmine erinevus hindajate skoorist: 12. klass

Hindamise aspekt	M – H	H1 – H2
ÕIGEKIRI JA VORMISTUS (0–9 p)	0,99	1,34
Kirjavahemärgistus (0–3 p)	0,52	0,70
Õigekiri ja vormimoodustus (0–3 p)	0,56	0,60
Liigendus ja vormistus (0–3 p)	0,45	0,60
LAUSESTUS JA SÕNAVALIK (0–6 p)	0,78	1,01
Sõnavalik (0–3 p)	0,39	0,52
Lausemoodustus (ühildumine, sõnajärg, rektsioon) (0–3 p)	0,48	0,6



Vahejäreldusi statistilise keeletöötuse põhjal

- Masina hinnang erineb inimeste keskmisest vähem, kui erinevad inimeste hinnangud teineteisest
- Õigekirja ja vormistuse aspektis oli täpsem 9. klassi kirjandite automaathindamine, lausestuse ja sõnavaliku puhul erinevust ei ilmnenu
- Vajalik on täpsem vealiigitus (sõnaasendused, kobarparandused)

Hindamine suurte keelemudelitega (LLM)

- Eeltreenitud mudelid, kontekstiõpe (*in context learning, zero-shot learning*) ehk ei vaja eraldi treenimist iga tunnuse/ülesande jaoks
- Masinat "juhendatakse" sama hindamisskeemiga mida inimhindajad kasutavad
- Katse: LLMid järgivad järgivad hindamisjuhendeid hästi; odav ja kiire
- Masinat ei usuta pimesi: süstemaatiline võrdlus inimhindajatega
- Samas väljakutse: inimhindajate vahel ka suur variatsioon

Hindamine suure keelemudeliga

- Näitena GPT-4o (OpenAI) ja Gemini (Google) mudelid + 9. klassi katsekirjandid
- Üks sisend = juhend ühe kategooria hindamiseks + üks kirjand (ehk iga kategooria siin hinnatud eraldi)
- Võib ka teha: kogu hindamisskeem + üks kirjand



Ühe prompti komponendid:

Your task is to grade this Estonian 9th grade exam essay kirjand ...

For reference, the student was asked to do the following:

"Tutvuda alustekstidega ja ...

Use the following criteria to assess how many points to award...

3 if: sõnavalik on isikupärane ja rikkalik; sõnavalik sobib kirjakeelsesse teksti; võib esineda üksik sõnastusraskus või sõnastusviga.

2 if: sõnavalik mitmekesine, esineb üksikuid sõnakordusi ...

First explain your reasoning of relevant rubric criteria VERY briefly using keywords or phrases

Essay text begins here:

""""

Your task is to grade this Estonian 9th grade exam essay kirjand in ""triple quotes"" below, using the provided grading rubric. For reference, the student was asked to do the following: "Tutvuda alustekstidega ja kirjutada nende põhjal 200-sõnaline kirjand, kus mõtiskleda teismeliste tootumisharjumuste üle. Pealkirjastada kirjand. Kirjutada sissejuhatus, kus püstitada probleem ja sõnastada enda põhiseisukoha. Toetada oma põhiseisukohta vähemalt kahe teemaarenduslõiguga. Igas lõigus esitada alaväide, milles kasutada vähemalt ühte alusteksti näidet. Lisada ka enda näiteid. Kokkuvõttes esitada peamised järeldused." Here you will ONLY grade this aspect: Sõnavalik. Grade this on a scale of 0 to 3 points. Keep in mind this is just a text by an Estonian 9th grader. Use the following criteria to assess how many points to award, following this principle: if all the criteria are fulfilled then give 3 points but if something is lacking then lower points as described in this rubric:

3 if: sõnavalik on isikupärane ja rikkalik; sõnavalik sobib kirjakeelsesse teksti; võib esineda üksik sõnastusraskus või sõnastusviga.

2 if: sõnavalik mitmekesine, esineb üksikuid sõnakordusi; aga sõnavalik sobib kirjakeelsesse teksti; esineb mõnesid sõnastusraskusi.

1 if: sõnavalik ühekülgne, esineb palju sõnakordusi; sõnavalik sobib suuremalt jaolt kirjakeelsesse teksti; esineb palju sõnastusraskusi.

0 if: sõnavalik ei sobi kirjakeelsesse teksti ja sõnastusraskuste tõttu tekst arusaamatu.

Note that sõnakordus here means repeating the same content word close by where a synonym would flow better, for example toitu in "Õpilased lähevad tihit poodi toitu ostma. Nad ostavad sealt ainult kommi kuna pood ei paku korralikku toitu."

First explain your reasoning of relevant rubric criteria VERY briefly using keywords or phrases in Estonian in this comma-separated format:

"Seletus: asjaolu: kirjeldus, asjaolu: kirjeldus, Hinne: N" ending with N points suitable grade. Express relevant asjaolud like "probleemile: vastatud", "laused: arusaadavad kuid esineb eksimusi", "õigekiri: palju puudusi; trüki vigu: ei esine"). Do NOT provide text examples in your reasoning.

Essay text begins here:

""""

Noorte tootumine ja aktiivsus

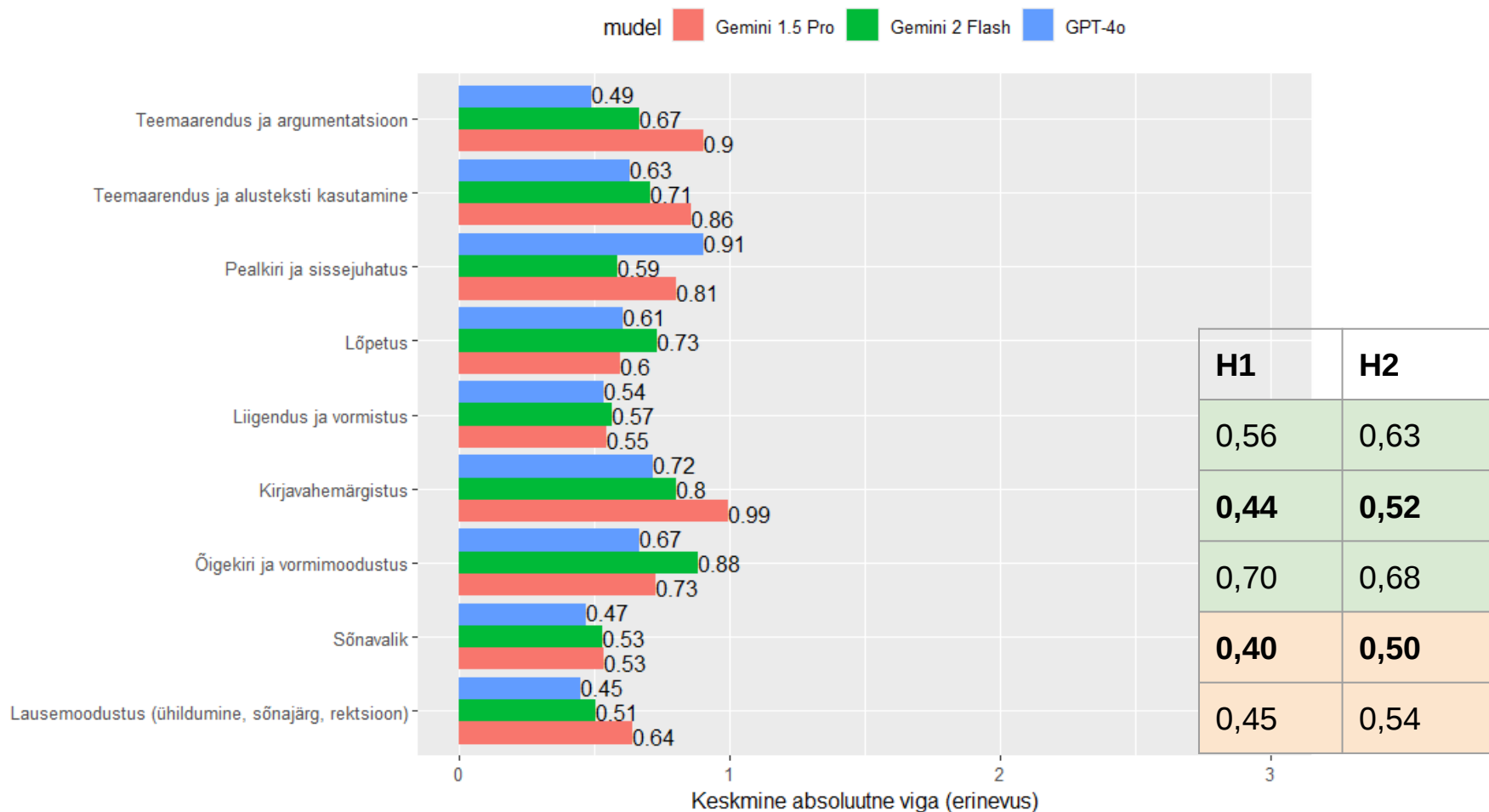
Toitumisest räägitakse igal pool.

.....

Nende vahel tuleb leida tasakaalu, et saada endale rahuldav kehakuju ja püsida terve, sest tasakaal on peaaegu.

""""

GPT-4o ja Gemini, erinevus võrreldes hindajate keskmisega



GPT-4o veamäär on kõige madalam

- Terve eksami skaala: 0...27 punkti
- Min erinevus 0, max 12, keskmine 2.62 ehk 9.7%
- Kui hindajate keskmise asemel võtta *lähim* hinne, siis keskmine absoluutne erinevus on 0.38 (0..27 skaalal!)
- Ja: **100%** GPT hinnetest jäi kahe hindaja vahele!
- Kulu: ~10M tokenit ~ \$25-30 terve eksami hindamine

Järeldused ja diskussioon

- Keeletehnoloogilised tööriistad pakuvad võimalust suures koguses kirjatöid kiiresti hinnata, ja ka päris hea täpsusega
- Tööriistade optimaalsus: keelemudelid sisulistele kriteeriumitele, statistiline keeletöötlus grammatika ja ortograafia jaoks
- Masinaid ei pea ega tohigi pimesi uskuda: pigem eksperdi võimestaja kui asendaja
- Edasi: suurema eksamimaterjali hindamine (sh 12. kl), tööriista prototüüpimine; paremad hindamisjuhendid ja õpetamismaterjalid