

მანქანური სწავლების ალგორითმები ტექსტის კითხვაადობის შეფასებისთვის

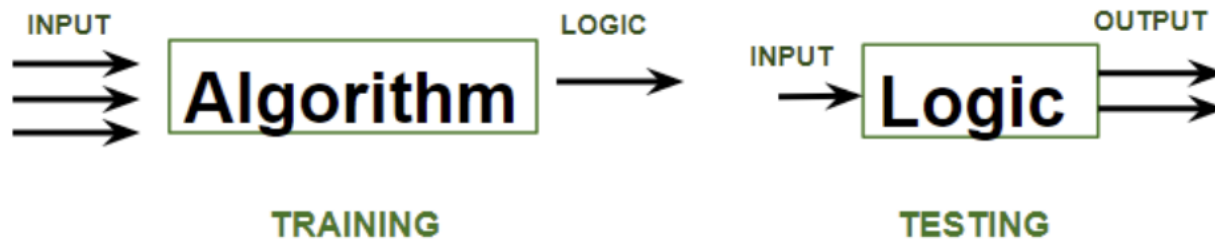
Machine Learning Algorithms for Assessing Text Readability

ასისტენტ-პროფესორი: ელზა ჯინჭარაძე

მაისი 2025

აბსტრაქტი

- რატომ არის ტექსტის კითხვადობა მნიშვნელოვანი (სფეროები): განათლება, ჯანდაცვა, ციფრული კონტენტი, საჯარო კომუნიკაცია
- მიზანი: როგორ აუმჯობესებს ML (Machine learning) ტექსტის კითხვადობის და გაანალიზების შეფასებას
- გადასვლა წესებზე დაფუძნებული მეთოდებიდან ჰიპოთეზური სისტემებზე



რა არის ტექსტის კითხვადობა?

- რამდენად მარტივია ტექსტის წაკითხვა და გაგება
- მნიშვნელოვანი მომენტები: სიტყვების და წინადადებების სიგრძე, გრამატიკა, ლექსიკა, კონტექსტი და სტრუქტურა

ტრადიციული კითხვადობის ფორმულები

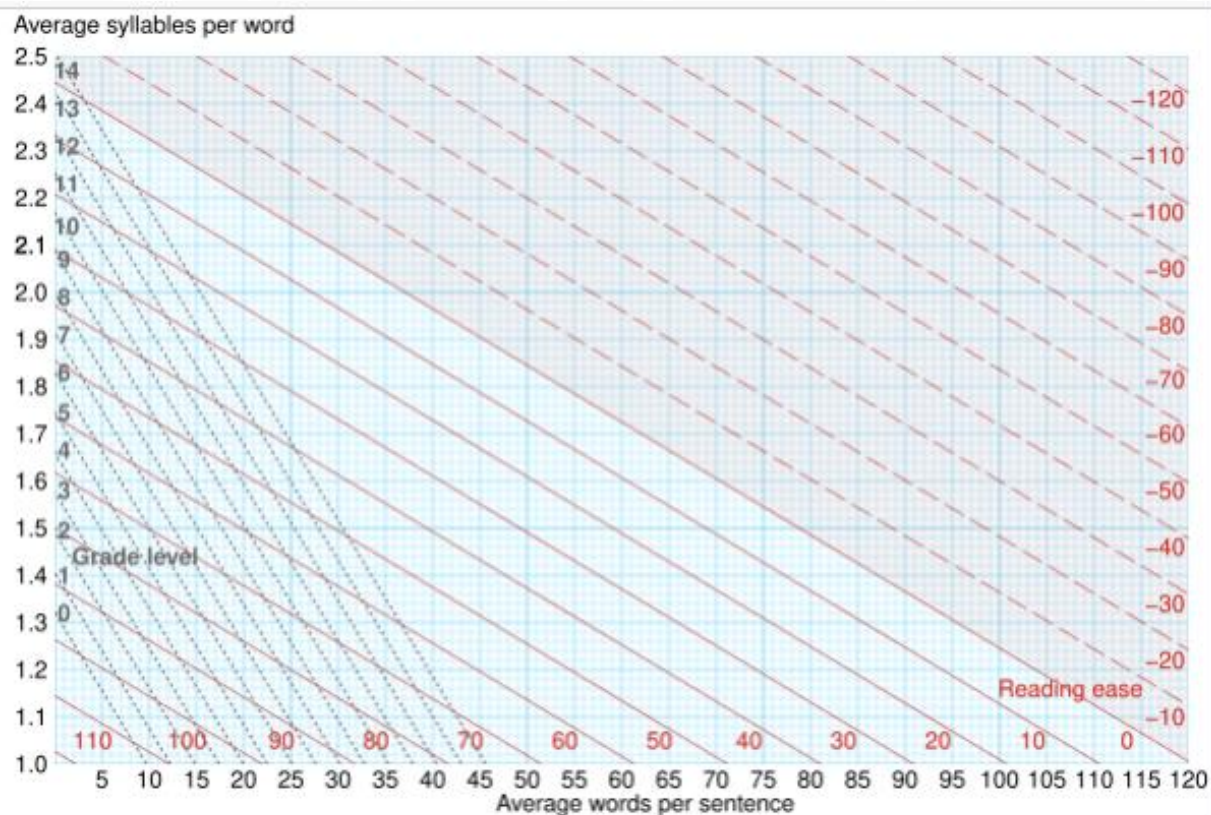
- მაგალითები: Flesch-Kincaid, Gunning Fog Index
- მოცემული სისტემები დაფუძნებულია ზედაპირულ მახასიათებლებზე
- შეზღუდვები: ითვალისწინებს მნიშვნელობას ან კონტექსტს, სფეროების სპეციფიკას

ლინგვისტიკაში GUNNING FOG INDEX არის ინგლისური წერის წაკითხვის ტესტი. ინდექსი აფასებს ფორმალური განათლების წლებს, რაც ადამიანს სჭირდება ტექსტის პირველი მოსმენით გასაგებად. მაგალითად, ინდექსი 12 მოითხოვს შეერთებული შტატების საშუალო სკოლის უფროსის (დაახლოებით 18 წლის) კითხვის დონეს. ტესტი 1952 წელს შეიმუშავა რობერტ განნინგმა, ამერიკელმა ბიზნესმენმა, რომელიც ჩართული იყო გაზეთებისა და სახელმძღვანელოების გამოცემაში.

$$0.4 \left[\left(\frac{\text{words}}{\text{sentences}} \right) + 100 \left(\frac{\text{complex words}}{\text{words}} \right) \right]$$

FLESCH-KINCAID წაკითხვის ტესტები არის წაკითხვის ტესტები, რომლებიც შექმნილია იმის დასადგენად, თუ რამდენად რთულია ინგლისურ ენაზე დაწერილი ტექსტის გაგება.

არსებობს ორი ტესტი: FLESCH READING-EASE და FLESCH-KINCAID GRADE LEVEL. მიუხედავად იმისა, რომ ისინი იყენებენ ერთსა და იმავე ძირითად ცვლადებს (სიტყვის სიგრძე და წინადადების სიგრძე), მათ აქვთ განსხვავებული შეფასების ფაქტორები.



რატომ გამოვიყენოთ მანქანური სწავლება?

სწავლობს რეალურ მონაცემებზე და
არა ფიქსირებულ წესებზე

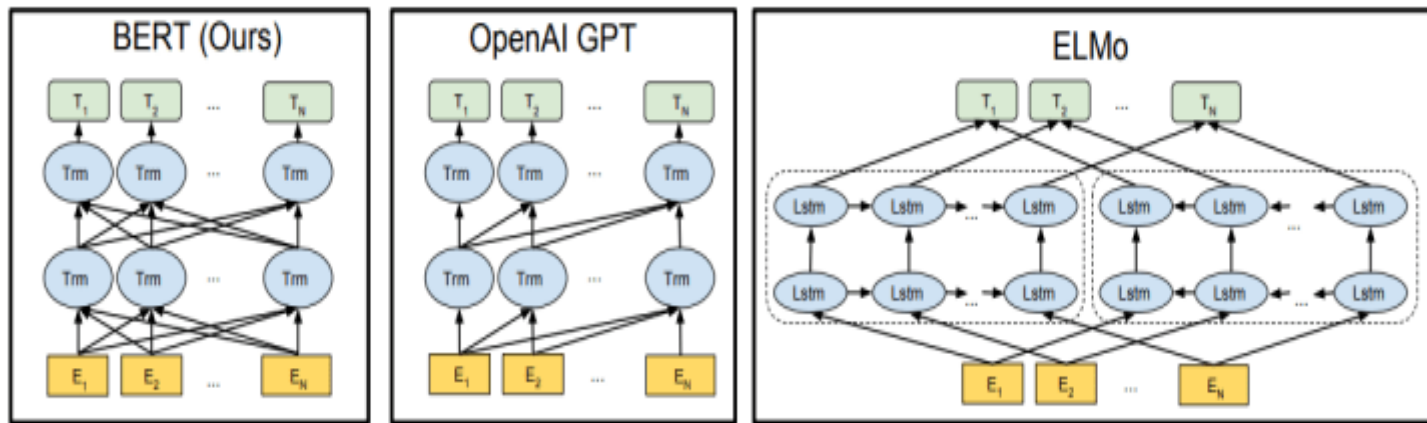
შეიცავს ღრმა ტექსტურ
მახასიათებლებს: სინტაქსი,
სემანტიკა, კონტექსტი

უფრო მოქნილი და მასშტაბირებადია

MACHINEL LEARNING (ML)

მიდგომები კითხვადობისთვის

- ზედამხედველობითი სწავლება: SVM, Random Forest
- ღრმა სწავლება: LSTM, ტრანსფორმერები - მაგ. BERT

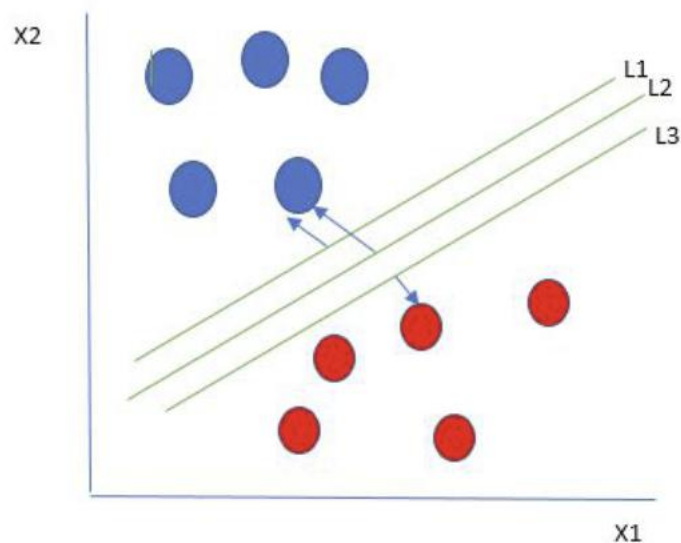


(Long Short-term Memory (LSTM) Is An Enhanced Version Of The Recurrent Neural Network (RNN) Designed By Hochreiter & Schmidhuber. Lstms Can Capture Long-term Dependencies In Sequential Data Making Them Ideal For Tasks Like Language Translation, Speech Recognition And Time Series Forecasting.)

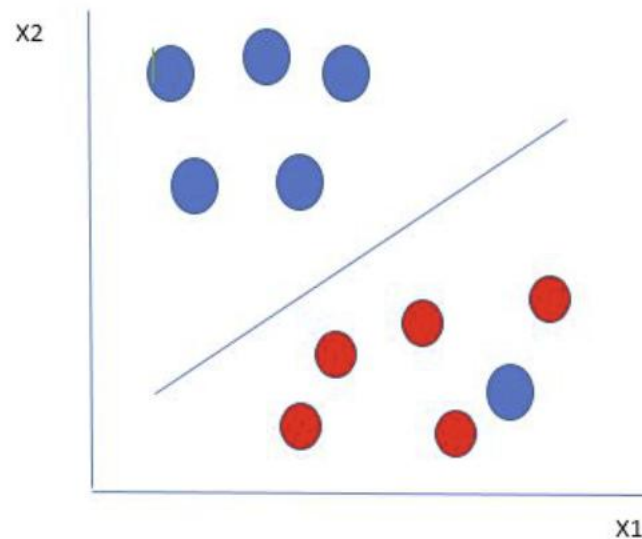
Bert (Bidirectional Encoder Representations From Transformers) Leverages A Transformer-based Neural Network To Understand And Generate Human-like Language.

SUPPORT VECTOR MACHINE (SVM) ALGORITHM

SVM არის ზედამხედველობითი მანქანური დასწავლის ალგორითმი, რომელიც გამოიყენება კლასიფიკაციისა და რეგრესიის ამოცანების შესასრულებლად. მას შეუძლია გაუმკლავდეს რეგრესიის პრობლემებს, SVM განსაკუთრებით კარგად შეეფერება კლასიფიკაციის ამოცანებს.



Multiple hyperplanes separate the data from two classes



Hyperplane which is the most optimized one

მახასიათებლები ML მოდელებისთვის

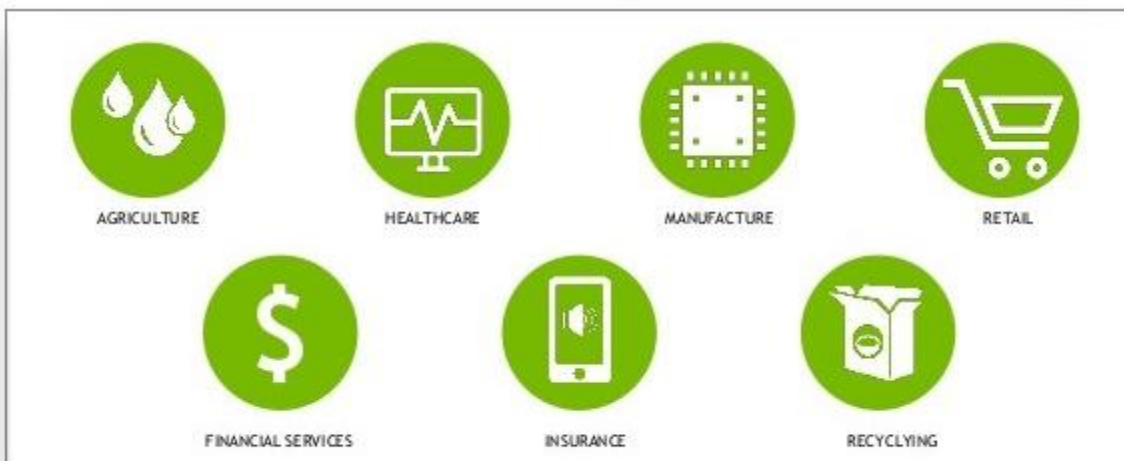
- ლექსიკური: სიტყვის სიგრძე, სიხშირე, მარცვლები
- სინტაქსური: გრამატიკა, წინადადების სტრუქტურა
- სემანტიკური: მნიშვნელობა, word embeddings
- დისკურსული: ტექსტის დინამიკა და კავშირი

სასწავლო მონაცემთა ბაზები

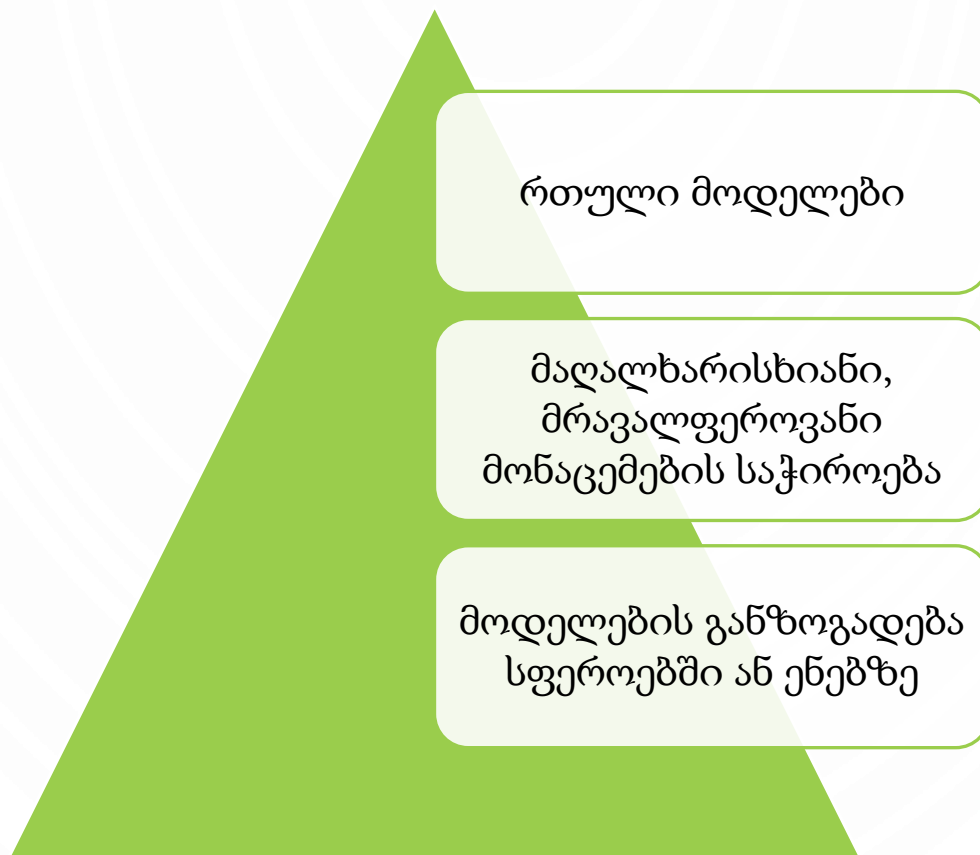
- **WeeBit:** საგანმანათლებლო ტექსტები სხვადასხვა ასაკისათვის
- **OneStopEnglish:** ტექსტები სამი სირთულის დონეზე
- **Newsela:** სიახლეები გადამუშავებული სხვადასხვა კითხვის დონეზე

რეალური გამოყენების მაგალითები

- განათლება: ადაპტირებადი სასწავლო ხელსაწყოები
- ჯანდაცვა: მარტივი ინსტრუქციები პაციენტებისთვის
- მედია და გამომცემლობა: შესაბამისი კონტენტი სხვადასხვა აუდიტორიისთვის
- მთავრობა: მკაფიო საჯარო დოკუმენტები



გამოწვევები და შეზღუდვები



სამომავლოდ კვლევის მიმართულებები

- ✓ მრავალენოვანი კითხვადობის ხელსაწყოები
- ✓ AI-ის ახსნადობა კითხვადობის შეფასებაში
- ✓ პერსონალიზაცია: ტექსტის მორგება ინდივიდზე

დასკვნა

- ML აუმჯობესებს კითხვადობის შეფასების სიზუსტესა და მოქნილობას
- აძლევს ღრმა გაგებას ტრადიციულ მეთოდებზე მეტად
- პერსონალიზებულია



გმადლობთ ყურადღებისთვის!