Rīgas 64. Vidusskola

**Sentimenta novērtēšana konkrēta YouTube video komentāru sadaļai**

Zinātniski pētnieciskais darbs: Datorzinātņu un datorlingvistikas sadaļā

**Darba autors:**

Rīgas 64. vidusskolas 12. klases skolnieks

Kristers Laganovskis

**Darba vadītājs:**

Rīgas 64. vidusskolas programmēšanas skolotājs

Edvards Bukovskis

Rīga 2023

**Anotācija**

Zinātniski pētniecisko darbu: “Sentimenta novērtēšana YouTube komentāru sadaļai” Izstrādājis autors Rīgas 64. vidusskolas 12.DIT klases skolēns – Kristers Laganovskis. Darba Vadītājs Rīgas 64. vidusskolas programmēšanas skolotājs - Edvards Bukovskis.

Zinātniski pētnieciskais darbs tika veikts ar mērķi, izpētīt kā ar dabiskās valodas apstrādi (NLP), iegūt sentimenta analīzi. Kā, tiek veikta efektīva datu ievākšana. Kā tiek veikta efektīva datu apstrāde. Kā arī cik precīzi ir iespējams novērtēt komentāru sadaļu balsoties uz mākslīgo intelektu un programmēšanas palīdzību. Zinātniski pētnieciskajā darbā balstoties uz analizēto un apkopoto informāciju par dabiskās valodas apstrādi, datu ievākšanu, datu apstrādi, tika izveidots pārlūka paplašinājums, ar kura palīdzību spēja analizēt sentimentu kādam konkrētam YouTube video komentāru sadaļu. Teorētiskajā daļa tika detalizēti izpētīti dabiskās valodas apstrādes (NLP) pamati, teksta sentimenta analizēšanai. Tās pielietošana sociālajā un ētiskajā kontekstā, kā arī tās nozīme mūsdienu un vēsturiskajā kontekstā. Tika apskatīts kas ir VADER modelis, kas ir RoBRTTa modelis un kā tie atšķiras. Kā, tiek veikta efektīva datu ievākšana un tā principi. Kā arī datu apstrādes principi, lai nodrošinātu pēc iespējas precīzāku sentimenta novērtējumu. Praktiskā daļā tika izstrādāts pārlūka paplašinājums, kas veidots uz VEDER un RoBERTa modeļu bāzes, šis paplašinājums sniedz iespēju vizualizēt statistisku par kāda konkrēta YouTube video komentāru sadaļas sentimentu, sniedzot vērtīgu perspektīvu par to, kāda ir publikas reakcija uz konkrēto saturu video. Paplašinājums nodrošina iespēju efektīvi filtrēt komentārus pēc to sentimenta, padarot analīzes procesu vēl efektīvāku.

Atslēgas Vārdi: Datu ievākšana, datu apstrāde, Komentāri, sentimenta analīze, pārlūka paplašinājums, dabiskās valodas apstrāde (NLP), mašīnmācīšanās, VADER, RoBERTa.

**Abstract**

Scientific Research Work: "Sentiment Analysis of YouTube Comments Section"

Authored by Kristers Laganovskis, a 12th-grade student at Riga 64th Secondary School. Supervised by Edvards Bukovskis, programming teacher at Riga 64th Secondary School.

The scientific research work aimed to explore sentiment analysis using natural language processing (NLP), focusing on effective data collection and processing. It also assessed the accuracy of evaluating the comments section based on artificial intelligence and programming assistance. Drawing from analyzed and compiled information on natural language processing, data collection, and processing, a browser extension was developed to analyze sentiment in specific YouTube video comment sections. The theoretical part extensively examined the fundamentals of natural language processing (NLP) and text sentiment analysis, considering its application in social and ethical contexts, as well as its significance in modern and historical contexts. The VADER model and the RoBERTa model were discussed, highlighting their differences. Effective data collection and its principles were examined, along with data processing principles to ensure the most accurate sentiment assessment.

In the practical part, a browser extension was developed based on the VADER and RoBERTa models. This extension allows for visualizing statistical data regarding the sentiment of a specific YouTube video's comment section, providing valuable insight into the audience's reaction to the content. The extension enables efficient filtering of comments by sentiment, making the analysis process even more effective.

Keywords: Data collection, data processing, comments, sentiment analysis, browser extension, natural language processing (NLP), machine learning, VADER, RoBERTa.

**Saturs**

Anotācija…......................................................................................................................................2

Abstract...........................................................................................................................................2

Ievads…………………………………………………………………………………………………………………………………..…..4

Secinājumi………………………………………………………………………………………………………………….……………..0

Izmantotie informācijas avoti…………………………………………………………………………………………………….0

Pielikumi……………………………………………………………………………………………………………………………………0

**Ievads**

Mūsdienās informācijas klāsts ir liels un apjomīgs, taču tas var radīt savas neērtības gan kā satura patērgātājiem, gan kā satura veidotājiem. Problēma ir meklējama tajā, ka cilvēki bieži vien pārāk vērš lielu uzticību satura veidotājam, kas ne vienmēr var būt labi, jo satura veidotājs var speciāli apslēpt un manipulēt ar to patērētājiem. Tāpēc ir ļoti svarīgi zināt, ko citi cilvēki domā par noskatīto video saturu. Salīdzinoši nesen 2021. gada. Novembrī YouTube noņēma iespēju redzēt “nepatīk” balsu skaitu, kas ietekmēja patērētājus negatīvi. Šāds lēmums veicināja nekorektu novērtējumu par video saturu, jo neuzrādoties “nepatīk” balsu skaitam, ir grūti spriest, cik cilvēki ir negatīvi novērtējuši video saturu. Tāpēc viens no veidiem, kā noskaidrot, ko cilvēki domā par video saturu ir vērsties pie komentāru sadaļas, taču tad atkal rodas problēma. Ja tie ir desmit komentāri, tas nerada lielas neērtības, bet, ja tie ir simti vai tūkstoši, tas noteikti rada neērtības, jo šādu apjomu izskatīt aizņems ļoti ilgu laiku.

Šajā darbā tiks pētīta kāda YouTube video komentāru sadaļa, kā ar valodas apstrādes (NLP) modeļiem spēj nolasīt komentāru sentimentu, kādas metodes tiek pielietotas, lai iegūtu nepieciešamos datus analīzei, kā arī kādā veidā dati ir jāapstrādā, lai veiksmīgi iegūtu sentimenta analīzi. Kā arī cik precīzi spēj nolasīt komentāru sentimentu balstoties uz mākslīgo intelektu. Tiks veidots arī paplašinājums ar kura palīdzību ikviens varēs analizēt komentārus, kā arī tos filtrēt pēc noteiktām kategorijām, kas ir noderīgi ne tikai satura uztvērējam, bet arī satura veidotājam attiecīgi ietaupot laiku filtrējot komentārus.

**Darba mērķis:** Izpētīt dabiskās valodas NLP principus kā arī salīdzināt VADER un RoBERTa modeļus un izveidot paplašinājumu, kas ļaus analizēt komentārus.

**Hipotēze:** Autoram izdosies veiksmīgi salīdzināt VADER un ROBERTa modeļus kā arī analizēt un interpretēt YouTube video komentārus, izmantojot NLP modeļus un programmēšanu.

**Darba uzdevumi:**

1. Izpētīt NLP modeli un kā to pielietot sentimenta novērtēšanai
2. Izpētīt atšķirības starp VADER un RoBERTa modeļiem.
3. Izpētīt, kā iegūt nepieciešamos datus no YouTube video komentāru sadaļas
4. Izpētīt, kā apstrādāt datus efektīvi, lai tie rezultētos precīzā novērtējumā
5. Izstrādāt paplašinājumu, kas ļauj ikvienam analizēt un filtrēt komentārus pēc noteiktām kategorijām.
6. Veikt secinājumus

**Izmantotās darba metodes:** Literatūras analīze, NLP (Dabiskās valodas apstrādes), VADER NLP modelis, RoBERTa NLP modelis. Datu iegūšana no YouTube video komentāru sadaļas. Statistikas un analīzes metodes, lai interpretētu rezultātus un veiktu secinājumus. Atvērtā pirmkoda pielietošana, “Python” programmēšanas valodas pielietošana radošās daļas veikšanai.

**Darba struktūra:** Darbs sastāv no x nodaļām, x apakšnodaļām, secinājumiem, izmantoto literatūras avotu saraksta un x pielikumiem. Darbā ir x attēli.