



**INTELEKTINIŲ METODŲ TYRIMAI NUOMONIŲ  
ANALIZEI DIDELIUOSE DUOMENŲ MASYVUOSE**

II ataskaitiniai metai

2015m. spalio mėn. 1d. – 2019m. rugsėjo mėn. 30d.

**Darbo vadovas:** prof. dr. Gintautas Garšva

# Tyrimo objektas, tikslai, planuojami gauti rezultatai

## **Tyrimo objektas:**

Intelektiniai metodai.

## **Tyrimo tikslas:**

Sukurti intelektinį metodą ar jų hibridą nuomonių analizei dideliuose duomenų masyvuose.

## **Tyrimo uždaviniai:**

- Apžvelgti esamus nuomonių/sentimentų analizės tyrimus ir identifikuoti problemą.
- Sukurti intelektinį metodą ar jo hibridą ir pagrįsti naujas savybes, požymius.
- Sudaryti programinį prototipo projektą, adaptuojant sukurtą metodą.

## **Planuojami rezultatai:**

Programinis prototipo projektas nuomonių analizei dideliuose duomenų masyvuose, adaptuojant sukurtą metodą.

# Ataskaitinių metų darbo planas

## **Studijų planas:**

Išlaikyti egzaminus:

- „Skaitinis intelektas“. Vertinimo komisija: doc. dr. V. Rudžionis, prof. dr. A. Lopata, prof. dr. G. Garšva;
- „Sistemų analizės technologijos“. Vertinimo komisija: prof. dr. S. Gudas, prof. dr. A. Lopata, doc. dr. V. Rudžionis;
- „Duomenų analizės strategijos ir sprendimų priėmimas“. Vertinimo komisija: prof. habil. dr. G. Dzemyda, dr. J. Bernatavičienė, dr. V. Medvedev;

## **Mokslinių tyrimų planas:**

- Mokslinių tyrimų disertacijos tema apžvalga ir analizė;
- Teorinis tyrimas;
- Tyrimo metodikos sudarymas.

## **Rezultatų pristatymo planas:**

- Dalyvavimas tarptautinėje mokslinėje konferencijoje Lietuvoje arba užsienyje.

## **Mokslinių publikacijų planas:**

- Planuojamas mokslinis straipsnis „SVM and Naive Bayes classification for sentiments analysis“.

# Ataskaita

## 2015–2017 m. m. išlaikyti egzaminai:

	<b>Dalyko pavadinimas</b>	<b>Kreditų skaičius ECTC</b>	<b>Atsiskaitymo data</b>	<b>Dalyko konsultantas</b>	<b>Įvertinimas</b>
1	Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika	9	2016.06 (2016.06.09)	Prof. dr. A. Čaplinskas	<b>8</b>
2	Skaitinis intelektas	7	2016.11 (2016.09.12)	Doc. dr. V. Rudžionis	<b>9</b>
3	Sistemų analizės technologijos	7	2017.03 (2016.10.20)	Prof. dr. S. Gudas	<b>9</b>
4	Duomenų analizės strategijos ir sprendimų priėmimas	7	2017.06 (2016.12.19)	Prof. habil. dr. G. Dzemyda dr. O. Kurasova dr. J. Bernatavičienė	<b>9</b>

Konstantinas Korovkinas, fiziniai mokslai, informatika (09P)



# Ataskaita

## 2015–2017 m. m. disertacijos rengimo planas:

Etapo pavadinimas	Planuota atlikti data	Pastabos
1 Mokslinių tyrimų disertacijos tema apžvalga ir analizė: 1. Intelektinių metodų dideliuose duomenų masyvuose tyrimų apžvalga. 2. Nuomonių/sentimentų analizės tyrimų apžvalga. 3. Problemų identifikavimas. 4. Mokslinė hipotezė.	2016.09	Įvykdyta. Studijų eigoje bus pildoma.
2 2.1. Teorinis tyrimas: 1. Siūlomas intelektinis metodas ar jų hibridas, naujų savybių, požymių pagrindimas. 2. Kuriamo metodo palyginimas su jau egzistuojančiais metodais.	2016.12	Dalinai įvykdyta. Studijų eigoje bus pildoma.
3 2.2. Tyrimo metodikos sudarymas: 1. Tyrimo metodikos parinkimas. 2. Teorinio ir eksperimentinio tyrimų planavimas, sukurtajam intelektiniam metodui.	2017.09	Įvykdyta. Studijų eigoje bus pildoma.

# Ataskaita

## Dalyvavimas konferencijose 2015–2017 m. m.:

- nėra

## Publikacijos 2015–2017 m. m.:



The screenshot shows the submission interface for the Baltic Journal of Modern Computing. At the top, there is a navigation bar with the journal's logo, name, and links for Home, Journal, and Help. On the right, the user's role is 'Author', and the user name is 'konstantinask'. Below the navigation bar, the section is titled 'New submissions'. A table lists the submission details for ID BJMC1710-002, which is an Article by Konstantinas Korovkinas, Paulius Danėnas, and Gintautas Garšva. The article title is 'SVM and Naive Bayes classification ensemble method for sentiments analysis', submitted on 2017-10-13. Below the table, there is a 'Start new submission' button and a link to 'Back to Author menu'.

ID	Authors	Submitted	Round	Decision	Revision due
Type	Title			Decision date	Revision uploaded
BJMC1710-002	Konstantinas Korovkinas, Paulius Danėnas, Gintautas Garšva	2017-10-13		-	-
Article	SVM and Naive Bayes classification ensemble method for sentiments analysis			-	-

## Kita veikla 2015–2017 m. m.:

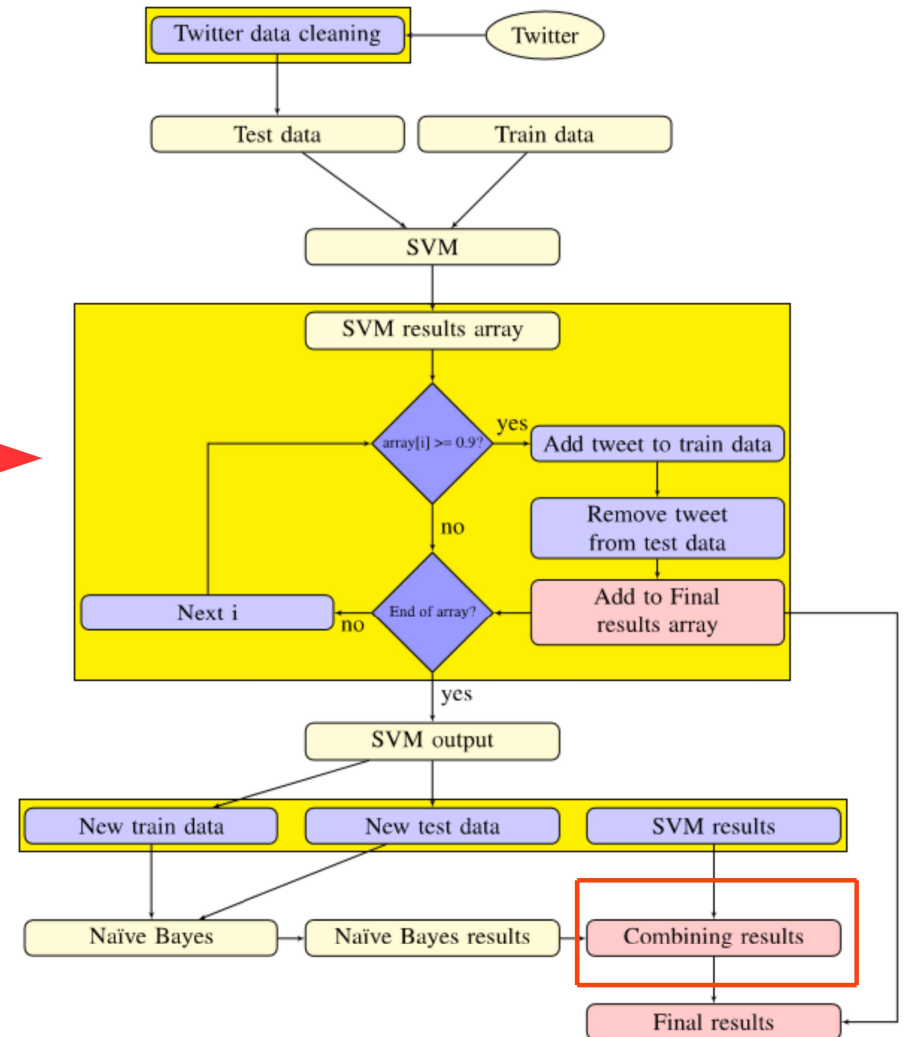
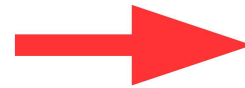
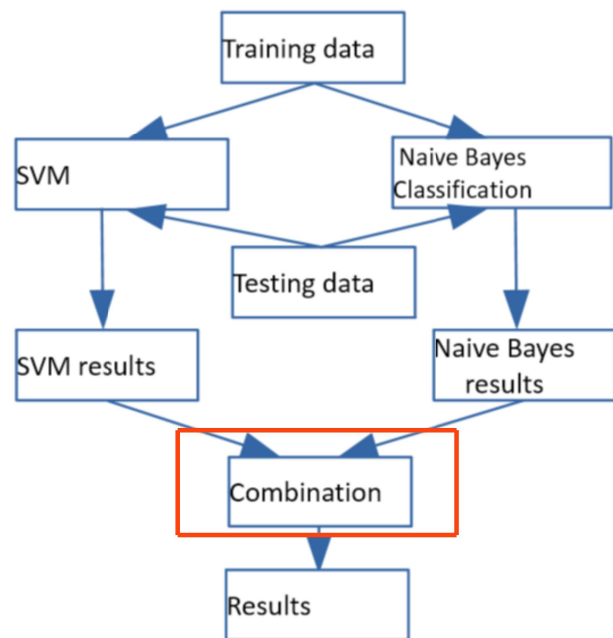
**2016m.** Bakalauro kursinio darbo vadovas (darbas įvertintas 8) ir baigiamojo darbo vadovas (darbas įvertintas 10).

**2017m.** Bakalauro kursinio darbo vadovas (darbas įvertintas 7) ir baigiamojo darbo vadovas (darbas įvertintas 9).

Konstantinas Korovkinas, fiziniai mokslai, informatika (09P)

# Ataskaita

## Siūlomas intelektinis metodas:



# Ataskaita

## Algoritmo žingsniai:

Algorithm for words. Let us denote the probability of word selection as  $p$ , and the threshold for its selection as  $th_1$ . Then the algorithm for  $SVM_{resultVALUE}$  is as follows:

1. Split sentence into words.
2. Select word if  $p \geq th_1$
3. Sum probabilities of selected words (if the word is “negative” the probability value is inverted as  $-p$ ).
4. Go to next sentence and start from step 1.

Naive Bayes classification is also performed as follows ( $Naive_{resultVALUE}$ ):

1. Split sentence into words.
2. Sum words if word is positive add 1, else minus 1.
3. Go to next sentence and start from step 1.

# Ataskaita

## Algoritmo žingsniai:

Algorithm for combining results if we split sentences into words:

1. Find results which are different between SVM and Naive Bayes Classification.
2. Find coefficient of difference, using our proposed formula:

$$difference = SVM_{resultVALUE} + \log_{10}(|Naive_{resultVALUE}|)$$

3. Find average of all coefficients of difference.
4. Check if coefficient of difference is smaller or equal average, then select  $SVM_{resultVALUE}$ ,  
else select  $Naive_{resultVALUE}$ .

# Ataskaita

## Algoritmo žingsniai:

Algorithm for sentences. This technique is applied on the whole sentence without splitting into words. Here we denote  $SVM_{resultVALUE}$  as the probability obtained after performing SVM classification,  $th_2$  as the threshold value for its selection and  $th_3$  as the threshold to select  $SVM_{resultVALUE}$ . Then the algorithm is as follows:

1. Find cases when  $SVM_{resultVALUE} < th_2$  and both SVM and Naive Bayes classifiers do not agree on their output for this case.
2. If  $|SVM_{resultVALUE}| < th_3$ , then select  $SVM_{resultVALUE}$ , else select  $Naive_{resultVALUE}$ .



# Ataskaita

## Rezultatai:

Exp No.	Training features	Training dataset	Testing features	Testing dataset	SVM results	Naive Bayes	Introduced method
1	words	6789	sentences	160	85.63	83.13	89,38
2	words	6789	sentences	1400	68,21	68,86	69,43
3	words +sentences	7629	sentences	560	75,71	52,68	74,46
4	words +sentences	7909	sentences	280	77,86	53,57	78,21

# Darbo planas kitai atestacijai

## Mokslinių tyrimų planas:

<b>Etaipo pavadinimas</b>	<b>Atlikimo terminas</b>
1 2.3. Eksperimentinis tyrimas: 1. Intelektinių metodų tyrimas nuomonių analizei dideliuose duomenų masyvuose, adaptuojant sukurtą metodą. 2. Programinis prototipo projektas, tyrimai, testavimas, galimi praktinės realizacijos variantai.	2018.09

Papildyti ir pataisyti pristatytas disertacijos dalis.

# Darbo planas kitai atestacijai

## **Mokslinių publikacijų planas:**

- Planuojamas mokslinis straipsnis “Support Vector Regression for stock market sentiment forecasting”.
- Planuojamas mokslinis straipsnis.

## **Dalyvavimas konferencijose, seminaruose, kitose doktorantų mobilumo veiklose:**

- Dalyvavimas Kauno fakulteto doktorantų tarpdisciplininiame seminare.
- Dalyvavimas konferencijoje “Information Society and University Studies” (IVUS) 2018
- Dalyvavimas tarptautinėje mokslinėje konferencijoje.

***Ačiū už dėmesį***