



Faculty of Electrical Engineering
and Informatics

Fake news detection in Slovak language using deep learning techniques

Bc. Klaudia Ivancová

klaudia.ivancova@student.tuke.sk

Ing. Martin Sarnovský, PhD.

martin.sarnovsky@tuke.sk

Ing. Viera Maslej-Krešňáková

viera.maslej.kresnakova@tuke.sk

Contents

- The definition of fake news
- Fake news detection
- Problem understanding
- Data understanding
- Data preparation
- Models training
- Conclusion

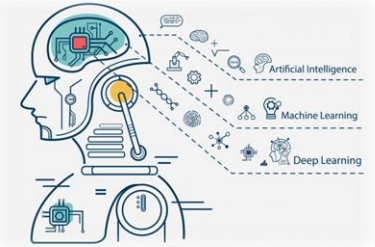
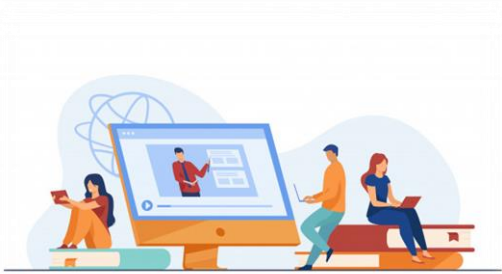


Fake news

- "Fake news is a news article that is intentionally and verifiably false."
- Fake news is not truthful and may be misleading or used to deceive the people.
- Using online technologies and platforms, fake news could be spread massively in a very short time.
- Early detection of the fake news is very important.
- The automated mechanisms are developed to detect the fake news.
- Machine learning and natural language processing (NLP) techniques are frequently utilized.
- Neural networks became very popular in recent years, also in fake news detection.



Fake news detection



- People are spending more and more of their lives in the online environment.

- There are many advantages like the ability to read, share, and publish information for all.
- Information exchange via the Internet takes less time and money.

- Although the quality of news published in unverified sources is much lower than in traditional portals, thanks to the mentioned advantages, people tend to search for such information more.

- Deep learning models became very popular to identify the fake news from the text.
- Advanced neural network models have been successfully and proved efficient in automated detection of the fake news.



Faculty of Electrical Engineering
and Informatics

Problem understanding

- This work aims to detect the fake news in the Slovak online space using deep learning methods.
- Our research task is focused on detecting fake news on a specific domain, and we have solved this task by creating models of neural networks.
- We focus on detection based on news content (text), without using any other data (images, authors or source info, etc.).
- We used recall, precision and F1 score to evaluate the models, and we wanted to achieve the best possible results.

Data understanding

- To collect the news articles in the Slovak language, we used the *MonAnt* platform.
- Individual articles can be obtained using *Postman* application.
- The final dataset contains articles from web sites *aktuality.sk*, *hnonline.sk*, *ta3.com*, *slobodnyvysielac.sk*, *zemavek.sk*, *magnificat.sk*, *panobcan.sk*, *hlavnespravy.sk* and *protiprudu.sk*.
- Target attribute is binary, **0** for True News or **1** for Fake News.
- Credibility of the news articles is set according to *konspiratori.sk*'s help, and manually by randomly selecting a few articles from each publisher, reading them in detail, and setting the labels manually.



Data preparation

- Removed unnecessary information from the collected articles and left only the body's content, i.e. the article's entire text and perex.
- Empty values - filled in up to 2,278 articles with text from the perex, if the article contained empty values for body and perex were removed.
- Balanced the target by down-sampling of the True News class records.
- The dataset consisted of 9 979 True News and 9 048 Fake News after preparation.



Faculty of Electrical Engineering and Informatics

name	url	body	perex
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/10/tady-to-je-a-bude-babisovo/	Na tento deň sme sa v Bratislave dočkali toho, čo sme si želali. Tady to je a bude Babišovo!	<p>Na tento deň sme sa v Bratislave dočkali toho, čo sme si želali. Tady to je a bude Babišovo!
redaktor maria miz	https://www.hlavnespravy.sk/mlady-futbalist...	Následkom toho došlo k vytrhnutiu dverí z ich osadenia. Žena. Vodič auta VW Golf prichádzal krátko pred 18.30 h smerom od B. Majakovskej ulice smerom na juhovýchod, z ktorého sa vyznačuje...	<p>Košice 28. januára 2018 (TASR/HSP/Foto:TASR)
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/10/rusk...	Zdroj: Rusko si chce natrvalo ponechať národnú základňu v sýr. Počas posledných dvoch dní moslavské úrady zadržali niekoľko ľudí...	<p>Na najaktuálnejšie spravodajstvo z domova, Bratislava 29. januára 2018 (TASR/HSP/Foto:TASR)
Martina Max	https://www.hlavnespravy.sk/galko-chcem-ab...	Na území Maďarska bude môcť podľa tohto ...	<p>Na území Maďarska bude môcť podľa tohto ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/10/mada...	Zdroj: Maďarský premiér Orbán predložil parlamentu návrh nove...	<p>Na území Maďarska bude môcť podľa tohto ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/11/euro...	Zdroj: Európske väznice a getá sú liahňou pre dýchadistov, va...	<p>Štúdia ICSR naznačila, že vznik a nárast...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/17/zabili-motorolu-velitela-doneckej-domobrany/	Zdroj: Zabili Motorolu, veliteľa doneckej domobrany - Svet - ...	<p>Arsen Pavlov zahynul pri výbuchu nálože ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/cins...	Napísal Igor Pejić pre SouthFront ; Editoval Desi TZONEV prek...	<p>Napísal Igor Pejić pre SouthFront ; Edit...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/11/priv...	Šokujúci prípad skupinového znásilnenia sa stal na švédskom o...	<p>Šokujúci prípad skupinového znásilnenia ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/11/puti...	Zdroj: Putin odložil návštevu Francúzska, kvôli diplomatickému...	<p>Podľa Kremľa sa 19. októbra má v Berlíne...
Martina Max	https://www.hlavnespravy.sk/danko-armensko...	Vzajomne vzťahy medzi krajinami uvoľnil za posledných desiat...	<p>Bratislava 29. januára 2018 (TASR/HSP/Fo...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/12/abde...	Zdroj: Abdeslam odmieta vypovedať, advokáti ho prestali zastu...	<p>Abdeslam je jediným žijúcim z hlavných p...
tomas zajaros	https://www.hlavnespravy.sk/na-poste-genera...	Na post manažera by teda mal zasadiť poslanec Marián Kéry. ...	<p>Bratislava 29. januára 2018 (TASR/HSP/Fo...
redaktor maria miz	https://www.hlavnespravy.sk/sudna-rada-sr-z...	Postup do druhého kola označil neúspešný Mazák za úspech. „V ...	<p>Bratislava 29. januára 2018 (TASR/HSP/Fo...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/12/fran...	Zdroj: Francúzii vyzývajú Európu k spojeniu. Chcú sa postaviť ...	<p>Starý kontinent by sa mal postaviť ameri...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/13/140...	Letecký útok na pohrebnu sálu, ktorý sa uskutočnil v súvislosti s ...	<p>Letecký útok na pohrebnu sálu, ktorý sa ...
TASR	https://www.hlavnespravy.sk/j-mihalcin-prac...	Urobme menšie územné celky, aby boli bližšie k ľuďom. Nesúhla...	<p>Vranov nad Topľou 2. novembra (TASR) – Prešovský región trpí vysoká miera nezamestnanosti. Má podl...
Martina Max	https://www.hlavnespravy.sk/firma-ccc-v-pra...	„Návrh na konkurz nás prekvapil, pretože so všetkými našimi v...	<p>Prakovce 23. marca 2015 (TASR/HSP/Foto:TASR)
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/13/amer...		nan
Martina Max	https://www.hlavnespravy.sk/prezident-proce...	Novelou sa má podľa ministerstva hospodárstva zabezpečiť posk...	<p>Bratislava 25. marca 2015 (TASR/HSP/Foto...
redaktor maria miz	https://www.hlavnespravy.sk/kabinet-by-mal...	Vláda však môže schváliť aj inú ako ministrom práce navrhovan...	<p>Bratislava 1. októbra 2019 (SITA/HSP/Fot...
Martina Max	https://www.hlavnespravy.sk/tomas-nagy-obdo...	„Tento vývoj priamo súvisí s tým, že sa severokórejskému reži...	<p>Bratislava 2. augusta 2017 (TASR/HSP/Fot...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/bara...	ROZHOVOR Politický analytik Ján Baránek vidí v novovzniknutej...	<p>ROZHOVOR Politický analytik Ján Baránek ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/cins...	Zdroj: ČÍNSKA VOJENSKÁ ANGAŽOVANOSŤ V AFRIKE WarsOnline Slo...	<p>Napísal Igor Pejić pre SouthFront ; Edit...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/belg...	Zdroj: Belgické Valónsko odmieta dohodu o voľnom obchode medz...	<p>Zákonodarcovia z Valónska, francúzsky h...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/irac...	Zdroj: Iracké sily pokročili v bitke o Músul, Kurdi zastavili...	<p>Boje o Músul sa začali v pondelok ráno, ...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/rusk...	Zdroj: Rusci a sýrske vzdušné sily zastavili nálety na mesto ...	<p>Odôvodom ich prerušenia je vytvorenie hum...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/18/tisi...	Zdroj: Chráma demonštranti vykrikovali protiturecké heslá a Tur...	<p>Chráma demonštranti vykrikovali protit...
slobodnyvysielac.sk	https://slobodnyvysielac.sk/2016/10/20/v-la...	Zdroj: V Las Vegas sa uskutočnil posledný televízny duel Clini...	<p>Las Vegas 20. októbra 2016 (TASR/HSP/Fot...
TASR	https://www.hlavnespravy.sk/bratislava-v-la...	Starosta Peter Šramko uviedol, že to, o čom poslankyňa Gabrie...	<p>Bratislava 2. novembra (TASR) – Ob...
redaktor maria dutkova	https://www.hlavnespravy.sk/rekonstrukcia-k...	Akadémia umení získala na rekonštrukciu kaštieľa 3,2 milióna ...	<p>Banská Bystrica 4. októbra 2019 (SITA/Hs...

- All punctuation marks, characters and hypertext links have been removed.
- The text has been tokenized, and also the stop words have been removed, then Word2Vec embeddings were used.
- Dataset was divided into the training set (11,410 samples), the validation set (3,803 samples), and the test set (3,804 samples).



body	label
na tenhle článok jsem se podstatě těšil celý letošní rok rád komentuji volby vždycky se sn...	1
najaktuálnější spravodajstvo domova ekonomiky financií tomu najkvalitnejší lifestyle na s...	1
bratislava januára vyňatie inšpekcie ministerstva vnútra mv sr podriadenosti ministra vnút...	0
na území maďarska bude môcť podľa tohto návrhu žiť prisťahovalec iba na základe individuál...	1
štúdia icsr naznačila že vznik nárast vplyvu skupiny islamský štát posilnil väzby medzi te...	1
arsen pavlov zahynul pri výbuchu nálože vo výťahu svojho bytu niekoľko ďalších ľudí utrpelo zranenia zdroj zabili motorolu veliteľa doneckej domobrany svet webnoviny sk	1
napísal igor pejić pre southfront editoval desi tzonev preklad pepe warsonline politické v...	1
šokujúci prípad skupinového znásilnenia sa stal na švédskom ostrove gotland meste visby že...	1
podľa kremľa sa októbra má berlína konať stretnutie tzv normandskom formáte najvyšších pre...	1
bratislava januára slovensko arménske vzťahy by sa mohli naďalej upevňovať spolupráca medz...	0
abdeslam je jediným žijúcim hlavných podozrivých účasti na sérii teroristických útokov kto...	1
starý kontinent by sa mal postaviť americkej praxi vymáhaniu obrovských platieb od európsk...	1
letecký útok na pohrebnú sálu ktorý sa odohral sobotu jemenskom hlavnom meste saná si vyži...	1
vranov nad topľou novembra vranovčan jozef mihalčin ročný konateľ súkromnej spoločnosti je...	0
prakovce marca vedenie rcc deklaruje že je naďalej odhodlané hľadať východisko pre firmu k...	0
bratislava marca novelu zákona investičnej pomoci podpísal prezident andrej kiska proces s...	0
bratislava októbra sociálni partneri sa na výške minimálnej mzdy na rok nedohodli tak sumu...	0
bratislava augusta severná kórea oznámila že uskutočnila už druhý úspešný test medzikontin...	0
rozhovor politický analytik ján baránek vidí novovznikutej strane richarda rybníčka toska...	1
napísal igor pejić pre southfront editoval desi tzonev preklad pepe warsonline politické v...	1
zákonodarcovia valónska francúzsky hovoriaceho regiónu belgicku piatok hlasovali proti uza...	1
boje mósul sa začali pondelok ráno pričom irackú armádu kurdske oddiely píšmarga podporila...	1
dôvodom ich prerušenia je vytvorenie humanitárnych koridorov zdroj ruskej sýrske vzdušné sily zastavili nálety na mesto aleppo	1
iraqi kurds protest turkish airstrikes in kurdish areas zhromaždení demonštranti vykrikova...	1
las vegas októbra dvaja hlavní kandidáti na prezidenta spojených štátov demokratka hillary...	1
banská bystrica októbra kompletná rekonštrukcia kaštieľa radvanských prilahlého parku by m...	0
politici vidia za dnešnými medializovanými obvineniami kontroverzného podnikateľa mariána ...	1
osvobození mosulu od povstalců může skončit totální katastrofou pro iráckou armádu koaličn...	1
slovo sa opäť prihlásil marián kočner urobil tak formou tlačovej konferencie ktorá včera v...	1
bratislava januára predseda ústavnoprávneho výboru nr sr róbert madej smer sd navrhuje za ...	0
damask na dnešok plánovaná prestávka bojoch aleppo je iba prvý krok aby mohli civilisti od...	1

Models training

MODEL 1		
Layer (type)	Output Shape	Parameters
Embedding	(None, 1600, 100)	18827400
Conv1D	(None, 1599, 100)	20100
GlobalMaxPooling	(None, 100)	0
Dense	(None, 256)	25856
Dense	(None, 1)	257
Total params:	18 873 613	
Trainable params:	18 873 613	
Non-trainable params:	0	

MODEL 2		
Layer (type)	Output Shape	Parameters
InputLayer	(None, 1600)	0
Embedding	(None, 1600, 100)	18887000
LSTM	(None, 128)	117248
Dropout	(None, 128)	0
Dense	(None, 128)	16512
Dense	(None, 1)	129
Total params:	19 020 889	
Trainable params:	19 020 889	
Non-trainable params:	0	

- All models were implemented in Python language using Tensorflow and Keras libraries.
- The experiments were conducted on a PC equipped with a 4-core Intel Xeon processor clocked at 4GHz and NVIDIA Tesla K40c GPU with 12 GB memory.

Model 1

	Precision	Recall	F1 score	Support
True News	0,91	0,95	0,93	2014
Fake News	0,94	0,89	0,92	1790
Accuracy			0,92	3804
Macro avg	0,93	0,92	0,92	3804
Weighted avg	0,92	0,92	0,92	3804

- Model 1 achieved an overall accuracy of 92,38%.
- We can see that a larger classification loss occurs in Fake News class, where up to 190 Fake News were classified as True News while 100 True News were classified as fake.

		Predicted	
		Fake News	True News
Actual	Fake News	True Positive (1600)	False Negative (190)
	True News	False Positive (100)	True Negative (1914)

Model 2

	Precision	Recall	F1 score	Support
True News	0,92	0,97	0,94	2014
Fake News	0,96	0,90	0,93	1790
Accuracy			0,94	3804
Macro avg	0,94	0,93	0,94	3804
Weighted avg	0,94	0,94	0,94	3804

- Model 2 achieved an overall accuracy of 93,56%.
- We can see that a larger classification loss occurred in the Fake News class, where 181 Fake News were classified as True News, while 64 True News were classified as fake.

		Predicted	
		Fake News	True News
Actual	Fake News	True Positive (1609)	False Negative (181)
	True News	False Positive (64)	True Negative (1950)

Conclusions

- The work described deep learning models for detection of fake news from the Slovak news articles.
- We approached this problem as a binary classification task - True news or Fake news.
- We trained 2 deep learning architectures - CNN and LSTM neural networks.
- We used only the text content of the articles and applied standard pre-processing methods.
- LSTM architecture could detect most of the false articles while producing fewer false negatives than the CNN approach.
- For future work, we will focus on the expansion and improvement of the dataset - particular topics and more data.





Faculty of Electrical Engineering
and Informatics

Thank you for your attention