

# Detekcia falošných správ na platformách sociálneho webu



Autor: Gabriela Demková, školiteľ: doc. Ing. Kristína Machová, PhD.

Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

## Motivácia

V dnešnej dobe technológií má k internetu prístup takmer tri a pol miliarda ľudí. Pri jeho zrode mala slúžiť pre dobro, a to s cieľom šíriť poznatky a vzdelanie, najskôr medzi akademikmi a neskôr medzi širokou verejnosťou. Keď sa neskôr začali objavovať sociálne siete ich cieľ bol podobný. Postupom času a ich rýchlym vývojom sa stali domovom falošných správ, klebiet, či nezmyslov, ktoré bohužiaľ používatelia ďalej zdieľajú bez overenia si správnosti. Každý deň je faloš a klamstvo šírené prostredníctvom sociálnych sietí z rôznych dôvodov, finančný zisk, získanie si priazne od najväčšieho počtu ľudí. A práve my, ľudia, tomu len dopomáhame.

## Riešenie problému

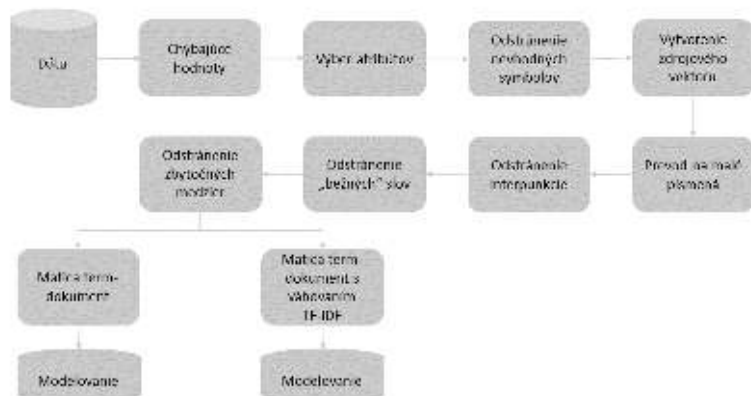
Hlavným cieľom je nájsť najpresnejší algoritmus strojového učenia, ktorý by mohol odhaliť tieto falošné správy. Na vyriešenie problému sme použili metodiku CRISP-DM. Atribúty boli upravené predspracovaním, čo je veľmi dôležité pri textových dátach. Hlavnou časťou je fáza modelovania, v ktorej sa menia vstupné údaje. Reprezentácia vstupu mala dve formy, maticu term dokument a maticu term dokument s váhovaním. Pri modelovaní sme na trénovaciu množinu aplikovali algoritmy strojového učenia, konkrétne *Naive Bayes*, *rozhodovacie stromy*, *náhodný les* a *podporné vektory*. Následne sme modely overovali na testovacej množine. Modely sme vyhodnocovali pomocou ukazovateľov binárnej klasifikácie, kde sme konkrétne sledovali: návratnosť, presnosť a F1 miera.

## Experimenty

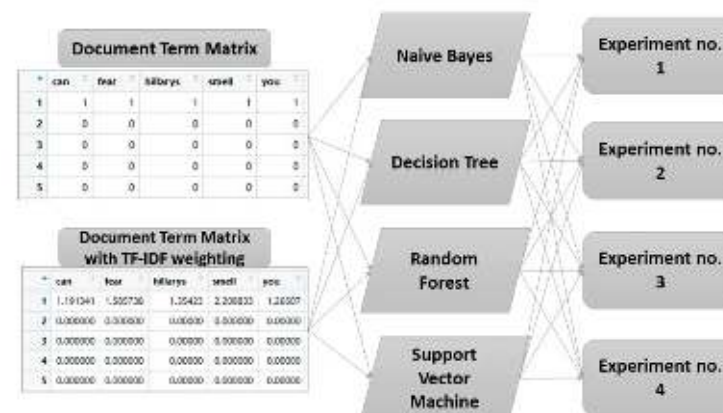
Attribute Headline	DTM			DTM with a TF-IDF weighting		
	Recall	Precision	F1 score	Recall	PRECISION	F1 score
NB	<b>0.792</b>	<b>0.829</b>	<b>0.810</b>	<b>0.820</b>	<b>0.826</b>	<b>0.823</b>
DT	1.00	0.543	0.704	0.200	0.898	0.327
RF	0.830	0.734	0.779	0.710	0.827	0.764
SVM	0.718	0.814	0.763	0.766	0.802	0.783

Attribute Body	DTM			DTM with a TF-IDF weighting		
	Recall	Precision	F1 score	Recall	PRECISION	F1 score
NB	0.910	0.818	0.862	0.880	0.936	0.907
DT	0.938	0.854	0.894	0.920	0.902	0.911
RF	<b>0.964</b>	<b>0.995</b>	<b>0.979</b>	<b>0.972</b>	<b>0.995</b>	<b>0.983</b>
SVM	0.570	0.938	0.709	0.916	0.978	0.946

## Predspracovanie dát



## Modelovanie



## Výsledky

Model Naive Bayes dosahoval najlepšie výsledky s menšou dátovou množinou na vstupe, jednalo sa o nadpisy príspevkov.

Model Random Forest dosahoval najlepšie výsledky s väčšou dátovou množinou na vstupe, ktorá pozostávala z textu príspevkov.