

## Predlog projekta

# Analiza rukopisa za otkrivanje osobina ličnosti primenom mašinskog učenja i algoritama za obradu slike

### Uvod

Analiza rukopisa ili grafologija je proučavanje fizičkih osobina i obrazaca rukopisa na osnovu koje se otkriva psihološko stanje pisca u vreme pisanja i njegov karakter. S obzirom na to da je rukopis odraz ličnosti svakog pojedinca, grafologija se zasniva na pretpostavci da između rukopisa i karaktera postoji pravilan psihofizički odnos. Prilikom analiziranja rukopisa, potrebno je uzeti u obzir nekoliko važnih činilaca: iskorišćenost prostora, prostor između reči, pravac, debljina linija, oblik i veličina slova, karakteristična slova, vezivanje slova, potpis itd. U zavisnosti od veština grafologa i dužine teksta, postupak analize rukopisa može oduzeti dosta vremena. Kako bi ovaj proces postao efikasniji i precizniji, razvijene su razne metode koje koriste algoritme za obradu slike i tehnike mašinskog učenja. U radu [1] izdvojeno je nekoliko karakteristika uzoraka rukopisa i izvršena je klasifikacija pisaca u 5 osobina ličnosti: energičan, ekstrovertan, introvertan, neuredan i optimističan. Histogram orijentisanog gradijenta (HOG) izdvaja karakteristike iz uzorka rukopisa pisca koje predstavljaju ulaz klasifikatora Support Vector Machine (SVM), pri čemu je izlaz osobina ličnosti pisca. Tačnost SVM klasifikatora je 80%. U radu [2] prikazano je poređenje različitih algoritama mašinskog učenja koji se koriste za analizu ličnosti i identifikaciju pisaca sa tačnošću algoritma.

Ovaj projekat bi se bavio analizom rukopisa primenom mašinskog učenja i algoritama za obradu slike, s ciljem otkrivanja osobina ličnosti pisca. Baza podataka koja bi se koristila je IAM Graph Database Repository [3]. Baza sadrži slike rukopisa na engleskom jeziku. U radu bi se izvršila komparacija mera uspešnosti tri klasifikatora: Naivni (engl. *Naive*) Bayes, stabla odlučivanja (engl. *Decision Trees*) i logistička regresija (engl. *Logistic Regression*). Takođe bi se rezultati rada uporedili sa rezultatima navedenih referentnih radova [1][2].

### Metod

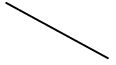


Ulazni podaci su slike rukopisa iz baze podataka IAM Graph Database Repository koja je dostupna na internetu.

Predloženi metod bi se sastojao iz 4 glavna modula: preprocesiranje slike, segmentacija, ekstrakcija karakteristika (engl. *features*) i klasifikacija.

U okviru preprocesiranja, vrši se odstranjivanje šuma na slici, normalizacija i kompresija podataka. Koriste se *grayscale* način zapisivanja slika. Predložene su sledeće metode: *histogram equalization*, gama transformacija, Furijeova transformacija, morfološke operacije (erozija i dilatacija), *thresholding* tehnike. Za uklanjanje šuma ispitivaće se *Median* i *Gaussian* filteri.

Segmentacija je proces deljenja slike na njene sastavne objekte ili regije. Segmentacija slika rukopisa podrazumeva segmentaciju linija, segmentaciju reči i segmentaciju slova.

Ekstrakcija karakteristika je tehnika smanjenja dimenzionalnosti ulaznih podataka i njihovo predstavljanje kao *feature vectors*. U ovom slučaju, izdvojeno je nekoliko važnih karakteristika rukopisa koje ukazuju na određene osobine ličnosti i prikazane su u sledećoj tabeli:

Svojstvo		Značenje	Svojstvo		Značenje
Veličina slova	Mala	Sposobnost koncentracije	Prekidi u pisanju	Povezanost	Objektivnost, racionalnost, logičnost
	Velika	Ambicioznost, dalekovidnost		Nepovezanost	Intuitivnost, osetljivost, nesigurnost
	Srednje veličine	Samopouzdanost, tradicionalnost, realnost			
	Promenljive veličine	Neodlučnost, brze promene raspoloženja			
Osnovna linija pisanja		Pesimizam, obeshrabrenost	Brzina pisanja	Brzo	Pamet, nekomunikativnost
		Optimizam, vera u budućnost		Sporo	Lenjost, nespretnost, neiskrenost
		Ujednačen temperament			
Pritisak na olovku	Blag	Neodlučnost, sposobnost brzog praštanja	Razmak između reči	Veoma širok	Odvojenost od stvarnosti
	Srednje jačine	Prosečan nivo emocionalnosti		Uzak	Štedljivost
	Jak	Dugotrajno pamćenje negativnih događaja		Širok	Mentalna spretnost, objektivnost
	Promenljiv	Promenljiva narav		Ujednačen	Ustrajnost, sistematično planiranje
Nagib	Udesno (BC)	Sklonost procenjivanju	Margine	Veliki razmak s leve strane	Hrabrost u suočavanju sa životom
	Ulevo (FA)	Opreznost, introvertnost		Veliki razmak s desne strane	Izbegavanje budućnosti, rezervisanost

	Vertikalno (AB)	Razum upravlja emocijama		Nema margina	Nesigurnost, pričljivost
	Promenljiv	Brze promene raspoloženja, nepredvidivost		Veliki razmak od početka papira	Formalnost, povučenost
	Udesno (CD)	Ekstrovertnost, orijentisanost ka budućnosti		Veliki razmak do kraja papira	Suzdržanost, površnost
	Udesno (DE)	Duboko proživljavanje svih situacija		Ujednačene	Samodisciplinovanost, samosvesnost
	Udesno (E+)	Velika ekspresivnost			

Tabela 1. Osobine rukopisa i objašnjenja pripadajućih karakternih osobina

Navedene karakteristike rukopisa izdvojiće se metodama koje su prikazane u radu [4].

Klasifikacija će se izvršiti na osnovu klasa prikazanih u tabeli 1. Ove klase predstavljaju izlazne podatke. Predložena je upotreba tri klasifikatora: Naivni Bayes (engl. *Naive*), stabla odlučivanja (engl. *Decision Trees*) i logistička regresija (engl. *Logistic Regression*).

## Metrike

Mere uspešnosti klasifikatora koje će se koristiti u ovom radu su tačnost (accuracy), preciznost (preccission), odziv (recall).

## Plan rada

Do zimskog seminara bih preciznije odredila metod za analizu rukopisa, kao i potrebne karakteristike (engl. *features*) rukopisa za određivanje osobina ličnosti. Takođe bih se upoznala sa navedenim tehnikama i bazom podataka koje koristim. Posle zimskog seminara bih započela implementaciju prva dva modula: preprocesiranje i segmentacija. Nakon toga bih implementirala ekstrakciju karakteristika slika rukopisa. Na letnjem seminaru bih završila klasifikaciju, testiranje i prikupljanje rezultata. Programski jezik koji ću koristiti je Python. Pisanje rada bih završila u periodu između letnjeg i jesenjeg seminara.

## Literatura

[1] Handwriting Analysis based on Histogram of Oriented Gradient for Predicting Personality traits using SVM. Aditya Chitlangia, G.Malathi

[2] A Survey: Machine Learning Approach for Personality Analysis and Writer Identification through Handwriting. Vishal Patil, Harsh Mathur.

[3] Baza podataka: <https://fki.tic.heia-fr.ch/databases/iam-graph-database>

[4] Sistem računarske vizije za analizu rukopisa, Ahmić Nermina.