

# Predviđanje pogrešaka izvornoga programskog kôda zasnovano na otkrivanju anomalija



Petar Afrić, mag. ing. comp.  
(petar.afric@fer.hr)

mentor: izv. prof. dr. sc. Marin Šilić,  
Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva



## 1. Uvod

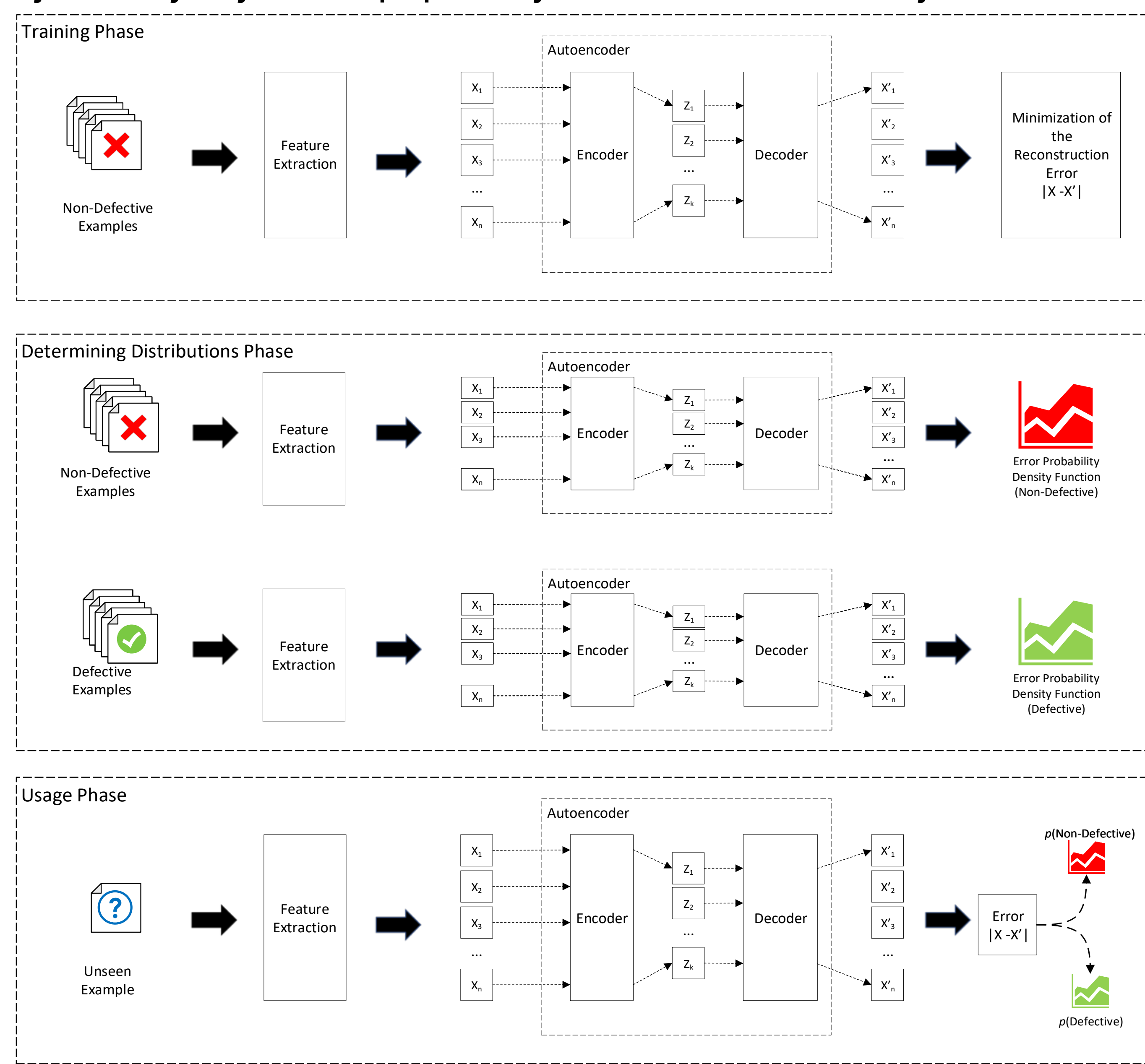
Količina napisanog programskog koda s godinama ubrzano raste. Napisani kod potrebno je održavati i nadograđivati, a pojava pogrešaka (engl. Bugs) u programskom kodu je neminovna. Programske pogreške mogu rezultirati velikim financijskim gubitcima, a u nekim slučajevima i ugrozom za život ljudi. Njihovo otkrivanje i otklanjanje spor je i skup proces. U cilju poboljšanja kvalitete napisanih programa te posljedično smanjenja financijskih gubitaka i povećanja sigurnosti, razvijaju se sustavi za automatsku identifikaciju programskih pogrešaka. Analizom svojstava programskog koda sustavi predviđa sklonost koda prema programskim pogreškama. Primjena sustava za automatsku identifikaciju pogrešaka u kodu pridonosi kvaliteti programskog koda, smanjuje financijske i sigurnosne rizike.

## 2. Opis problema

Podatkovni skupovi koji se koriste za razvoj metoda za otkrivanje pogrešaka u izvornom programskom kodu često su nebalansirani i sadrže malen broj primjera pogrešnog programskog koda. Zbog te činjenice cilj istraživanja je oblikovati i ostvariti učinkovitu metodu za automatsko otkrivanje programskih pogrešaka zasnovanu na ideji oblikovanja programskih pogrešaka kao anomalija s mogućnošću korištenja semantičkih značajki programa.

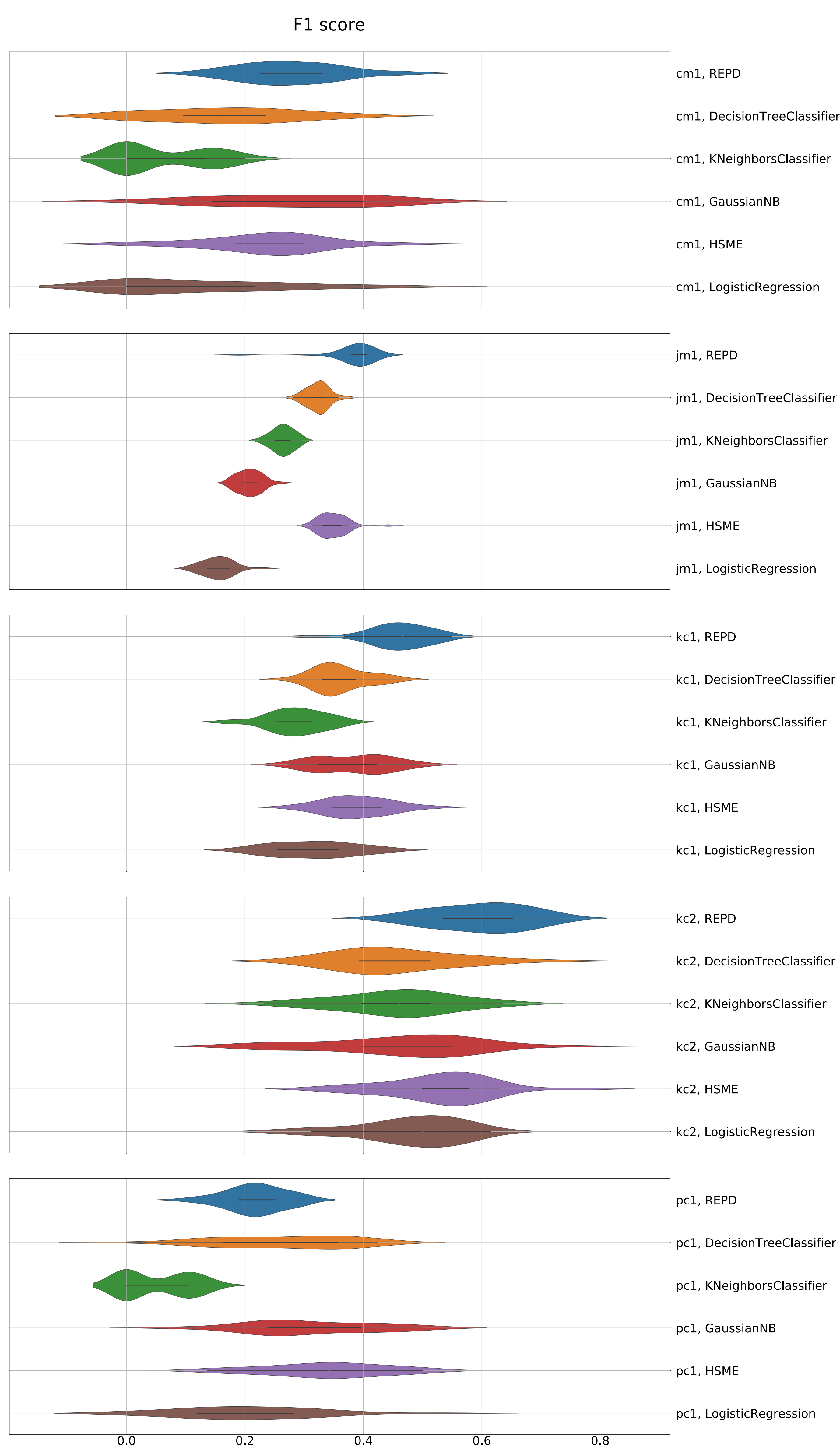
## 3. Metodologija

Prvi korak pri ostvarenju modela je ekstrakcija značajki programskog koda. Semantičke značajke moguće je izgraditi kodiranjem leksičkih tokena izvornog programskog koda. Sljedeći korak je samo treniranje modela. Predloženi model potrebno je prvo naučiti izgradnju reprezentacije ispravnog programskog koda. Potom se uče distribucije greška rekonstrukcije ispravnog i defektnog programskog koda. Konačno, model klasificira ne viđene primjerke temeljem njihove vjerojatnosti pripadanja naučenim distribucijama.



## 4. Rezultati

Dobiveni rezultati su statistički potvrđeni i pokazuju relevantnu superiornost predloženog pristupa nad klasičnim pristupima.



## 5. Zaključak

Istraživanje pokazuje da je problem moguće tretirati kao problem detekcije anomalija umjesto klasičnog pristupa binarne klasifikacije i time inherentno riješiti problem nebalansiranosti podatkovnih skupova.

