**Pregled postojeće relevantne literature**

Prilikom istraživanja radi boljeg upoznavanja sa tematikom ovog rada, pronađeno je nekoliko radova u kojima je obrađen sličan problem. Neki od njih su navedeni u nastavku teksta.

U radu [1] je izvršena predikcija rizika pokušaja samoubistava kroz vreme korišćenjem različitih tehnika mašinskog učenja. Korišćene su tradicionalne i novije tehnike mašinskog učenja, kako bi se omogućilo poređenje rezultata. Kao predstavnik tradicionalnog načina korišćena je logistička regresija, a algoritam *Random Forest* je upotrebljen u svrhu novije tehnologije mašinkog učenja. Predikcija je vršena na osnovu medicinskih podataka iz kliničkih zdravstvenih kartona u kojima su zabeleženi slučajevi povreda za koje se zna ili sumnja da su samonanešene. Takođe, u obzir su uzeti i zdravstveni kartoni u kojima nisu zabeleženi podaci o pokušajima samoubistva. Ovaj rad je pokazao uspešnost *Random Forest* algoritma za predikciju ovakvog tipa problema.

Rad [2] se bavi predikcijom ideja pojedinaca o samoubistvu na osnovu socio-demografskih, fizičkih i psiholoških obeležja. Predikcija se vrši upotrebom *Random Forest* algoritma. Od prvobitnih 47 obeležja, zapaženo je da je model obučen sa 39 karakteristika postigao najveću *Kappa* vrednost. Potom je eliminacijom atributa unazad (eng. backward selection) odbrano 15 obeležja za koje se pokazalo da postižu neznatno nižu *Kappa* vrednost. Kao neki od najbitnijih parametara su izdvojeni: depresija, nivo stresa u svakodnevnom životu, pol, starost. Kako bi se izbegao *overfitting* i povećala generalizacija modela, korišćena je 10-struka unakrsna validacija. Predočeni model je postigao tačnost 0,81. Iz prethodno izloženog, uočava se pogodnost korišćenja *Random Forest* algoritma za rešavanje problema ovog tipa. Takođe, zaključuje se i da je značajno koristiti unakrsnu validaciju kako bi model bio uopšteniji.

U radu [3] je razvijen model za predviđanje pokušaja samoubistava južnokorejskih adolescenata na osnovu stabla odlučivanja. Skup podataka predstavlja kompoziciju socio-demografskih i psiholoških podataka ispitanika (stepen depresije, delikvencija, samopouzdanje, optimizam) i podatka da li je ispitanik u prethodnih godinu dana pokušao da izvrši samoubistvo. Stablo odlučivanja sa sobom povlači opasnost od potencijalnog *overfitting*-a u slučajevima kada je prisutna velika količina podataka. U ovom radu je bilo svega 2754 ispitanika, te je ta mogućnost mala. Takođe, veliki broj promenljivih može da utiče na dubinu stabla, što otežava klasifikaciju. Kako je skup podataka za problem koji se opisuje u ovom radu mnogo veći od skupa podataka navedenog rada, nije preporučljivo da se koristi isti algoritam.

1. Colin G. Walsh, Jessica D. Ribeiro, Joseph C. Franklin (2017) *Predicting Risk of Suicide Attempts Over Time Through Machine Learning*
2. Seunghyong Ryu, Hyeongrae Lee, Dong-Kyun Lee, and Kyeongwoo Park (2018) *Use of a Machine Learning Algorithm to Predict Individuals with Suicide Ideation in the General Population*
3. Sung Man Bae, Seung A Lee, Seung-Hwan Lee (2015) *Prediction by data mining, of suicide attempts in Korean adolescents: a national study*