

Implementacija modela prediktivnog održavanja korištenjem binarne klasifikacije

Olivera Janković

ORAO a.d.

Bijeljina, BiH

olivera.jankovic@orao.aero

Sažetak—U fokusu ovoga rada je proces implementacije modela prediktivnog održavanja, obzirom na prirodu posmatranog problema – predikcije postojanja kvara avionskog motora u okviru određenog broja radnih ciklusa, u kontekstu modela binarne klasifikacije. Da bi se kreirali i u krajnjem uporedili dobijeni prediktivni modeli, kreirani su potrebni trening i testni setovi podataka na bazi seta podataka koji predstavlja simulaciju degradacije avionskog motora, korištenjem različitih tipova algoritama mašinskog učenja (primjenom algoritma logističke regresije, algoritma temeljenog na Bayesovim mrežama, algoritma stabla odlučivanja, algoritma k-najbližeg susjeda, implementacija algoritma vektora podrške i višeslojni perceptron).

Ključne riječi - prediktivni model održavanja, algoritmi mašinskog učenja, binarna klasifikacija;

I. UVOD

Troškovi zastoja opreme (mašina, uređaja, ...) u većini grana industrije mogu biti veoma značajni. Za industrije sa velikom i specijalizovanom opremom, kao što su transport, avijacija ili proizvodnja mogu da predstavljaju veliki dio budžeta. Dragocjenu i vrijednu opremu, transportna i ostala sredstva od poslovnog interesa, potrebno je čuvati u dobrom stanju kako ne bi došlo do zastoja, koji pak potencijalno može značiti velike i nenadoknadive gubitke za neku kompaniju[1]. Neka od istraživanja o efektivnosti upravljanja održavanjem govore da se jedna trećina troškova održavanja odnosi na nepotrebno i nepropisno obavljeno održavanje. Noviji, aktuelni načini poboljšanja efikasnosti i eliminisanja zastoja podržani su pored ostalog dostupnim podacima nastalim primjenom novih senzorskih i bežičnih tehnologija (M2M rješenja na primjer) [2]. Suočeni sa ubrzanim razvojem i primjerima efikasne primjene novih rješenja u praksi, proizvođači i pružaoci raznih servisnih usluga počinju da se okreću ka tendencijama prediktivnog održavanja koje su bazirane na preciznim podacima sa mašina/uređaja, raspoloživih u realnom vremenu, kako bi se utvrdilo stvarno činjenično stanje i na osnovu toga odredilo održavanje, što pored ostalog potencijalno nudi uštede troškova u odnosu na rutinsko ili vremenski bazirano preventivno održavanje, obzirom da se određene aktivnosti obavljaju u skladu sa potrebama odnosno stanjem mašine/uređaja..

Većina mašina/opreme je već i sada instrumentalizovana sa raznim vrstama senzora: temperaturnim, infracrvenim, akustičnim, sensorima vibracije i nivoa baterije, ...; sa ciljem da se monitorišu određeni uslovi i stanja što posljedično doprinosi dodatnom porastu u količini podataka (na već postojeće velike količine podataka različite prirode - vezano za proizvode, mašine, razne poslovne procese, održavanje, kontrolu kvaliteta, ...), tokom svakodnevnog poslovanja kompanije. Tome je svakako doprinio stalni napredak tehnologija računarskog hardvera, cjenovna prihvatljivost i sveprisutna dostupnost računara, opreme za prikupljanje podataka i medija za skladištenje. Međutim, samo obilje ovih podataka, bez moćnih alata za analizu, dovelo je do podacima bogatih ali informacijama siromašnih situacija u kojima ripozitorijumi podataka postaju arhive podataka, rjetko posjećene i korišćene. Prisutnost ovih podataka navela je istraživače da razviju algoritme koji automatski analiziraju podatke sa ciljem da se pronađu veze ili modeli koji mogu pomoći predvidjeti buduće neuspjehe. Ti algoritmi su uspostavili ono što se sada zove prediktivno održavanje. Jedna od opcija da se značajno poveća brzina kojom se pomenute količine podataka mogu pretvoriti u korisne informacije svakako je opcija data mininga [3][4], važno sredstvo za sticanje znanja iz dostupnih baza podataka, mogućnost da se predvide budući propusti otkrivanjem obrazaca ponašanja iz raspoloživih podataka. Predviđanje degradacije sistema, prije nego se dogodi, može dovesti do sprečavanja ili barem izbjegavanja loših posljedica

Poslovni problem ilustrovan u okviru rada vezan je u osnovi za predikciju problema uzrokovnog kvarom opreme, tako da se može odgovoriti na pitanje vezano za preostali radni vijek opreme, konkretnije u kontekstu korištenog seta podataka za predviđanje postojanja kvara (da li će se desiti ili ne) avionskog motora u nekom budućem periodu, u opsegu posmatranog, zadatog broja ciklusa. Dati problem se može formalno predstaviti kao problem binarne klasifikacije (klasa problema nadziranog učenja) te će se koristiti različiti tipovi algoritama mašinskog učenja da bi se kreirali i u krajnjem uporedili dobijeni prediktivni modeli koji se treniraju, obučavaju iz, u tu svrhu pripremljenog seta trening podataka. U okviru ovoga rada za ilustraciju implementacije prediktivnog modela u kontekstu modela binarne klasifikacije, (primjenom algoritma logističke regresije, algoritma temeljenog na

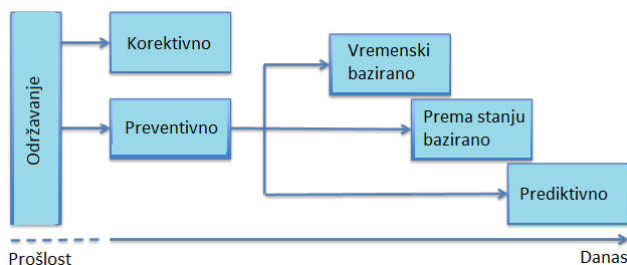
Bayesovim mrežama, algoritma stabla odlučivanja, algoritma k-najbližeg susjeda, implementacija algoritma vektora podrške i višeslojni perceptron - klasifikatori Weka open source softverskog alata), će se, kao polazna osnova, koristiti javno dostupan set podataka koji predstavlja simulaciju degradacije avionskog motora.

Važno je pomenuti da cilj ovog rada, obzirom na prirodu korištenog seta podataka, nije fokusiran na specifikaciju aktivnosti održavanja avionskih motora, već da se upoznaju i koriste nove tehnike i alati za pronalaženje korisnih znanja iz velikih količine sirovih podataka, koje mogu da pomognu da se poveća efikasnost sistema, što dovodi do poboljšanja u funkcionalnosti, koje se pak prenose na povećanje dostupnosti i samim tim u većoj produktivnosti. Namjera je da se ukaže na moguću pristup za predviđanje kvarova opreme, kroz prizmu problema koji se odnosi na avionski motor, uključujući u djelokrug održavanja dobro dokazane tehnike drugih oblasti nauke i tehnologije.

II. PREDIKTIVNO ODRŽAVANJE

Da bi se razumjelo šta je prediktivno održavanje, potrebno je prethodno razmotriti tradicionalne politike održavanja. Na Sl. 1 je prikazana evolucija strategija održavanja u vremenu [5]. Najranija tehnika (i najčešće korištena sve do sada) je korektivno održavanje (poznato i kao radom do otkaza (run-to-failure) ili reaktivno održavanje. U osnovi to je jednostavna i direktna procedura koja se sastoji od čekanja da se kvar dogodi da bi se zamjenio dio koji je defektan. Glavni nedostatak ovog pristupa uključuje rizičan i nepredvidiv ishod proizvodnje/poslovanja kao i najčešće visoke troškove neplaniranih održavanja.

Unapređenje dijagnostike industrijskih instrumenata, pored ostalog, doveo je do pojave nastanka vremenski vođene preventivne politike održavanja, kao što je preventivno održavanje [6] bazirano na rasporedu, na unaprijed definisanim vremenskim intervalima, gdje se dijelovi zamjenjuju prije nego se desio njihov kvar na primjer. Da bi se efikasno sprovedila ova periodična strategija, prethodno je potrebno napraviti neku vrstu istraživanja, procjenu da bi se pronašao optimalan raspored održavanja koji može da smanji troškove poslovanja i poveća dostupnost (planirani raspored uzima u obzir životni vijek opreme kao i broj neophodnog ljudstva i radnih sati; često samo na osnovu kvalitativnih informacija potaknutih iskustvom i inženjerskim obrazloženjima).



Slika 1. Evolucija politike održavanja kroz vrijeme

Razvoj inteligentnih senzora i uređaja utrlj su put za politiku održavanja baziranu na stanju. Radno stanje nekog

sistema se stalno prati pomoću namjenskih alata/opreme za praćenje i kada se degradacija otkrije sistem se mjenja. Ovaj metod povećava radni vijek i dostupnost komponenti i omogućava preventivne korektivne mjere (to zahtijeva povećano ulaganje u efikasnu opremu za monitoring, kao i obuku osoblja za održavanje.) Prediktivno održavanje [7] može da se predstavi kao mjera koje detektuju početak degradacije sistema, a samim tim i neposredni kvar, na taj način omogućavajući da se kontrolišu ili eliminišu uzroci kvara dovoljno rano da bi se izbjeglo bilo kakvo ozbiljno pogoršanje fizičkog stanja komponente. Prediktivno održavanje koristi neke parametre izmjerene u opremi da bi se dobio "osjećaj" kada je kvar istaknut. Ova vrsta održavanja ima za cilj da se osigura intervencija na opremi prije nego dođe do štetnih događaja.

III. MAŠINSKO UČENJE I PREDIKCIJA

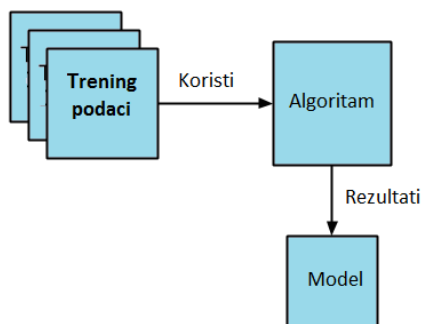
Mašinsko učenje, kao svojevrsan vid potrage za obrascima u podacima koji nas okružuju (putem algoritama i računarskih sistema), jedan je od načina koji može da se koristi za kreiranje prognostičkih/prediktivnih modela. Postoji mnogo algoritama mašinskog učenja na raspolaganju i generalno dva korisna načina njihovog kategorisanja. Prvi je grupisanje algoritama po stilu učenja. Drugi je grupisanje algoritama po sličnosti forme ili funkcije. Taksonomija ili način organizovanja algoritama mašinskog učenja po stilu je korisna jer tjera na razmišljanje o ulozi ulaznih podataka i procesa pripreme modela i odabira onog modela koji je najprikladniji za posmatrani problem kako bi se u krajnjem postigao najbolji rezultat. Kada se barata sa podacima za donošenje poslovnih odluka dva stila učenja, najčešće navođena su:

- Nadzirano učenje - svi ulazni podaci za trening imaju poznatu oznaku/labelu ili rezultat. Primjeri problema su klasifikacija i regresija.
- Nenadzirano učenje (podaci trening seta ne sadrže oznaku niti rezultat. Primjeri problema su klastering i učenje asocijativnih pravila (association rule learning).

Pored navedenih postoji i polu-nadzirano učenje (semisupervised learning) koje je u stvari kombinacija nadziranog i nenadziranog (često se koriste kada je trening skup mali).

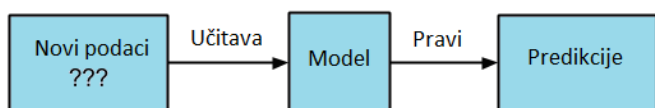
Algoritmi, važan i veliki dio mašinskog učenja, često su grupisani po sličnosti s obzirom na njihovu funkciju (kako rade, pri čemu treba imati na umu da postoje algoritmi koji se mogu lako uklopiti u više kategorija. Popularni algoritmi mašinskog učenja se tako grupišu na: algoritme regresije, na instanci bazirane algoritme, algoritme stabla odlučivanja, klastering algoritme, algoritme vještačkih neuronskih mreža, ensemble algoritme i ...

Prediktivno modeliranje je najčešći tip problema koji se rješava primjenom mašinskog učenja. Prilikom izgradnje prediktivnog modela potrebno je analizirati i povezati tri osnovna aspekta - podaci, model i predikcija, u kontekstu problema sa kojim se radi. U osnovi prediktivnog modeliranja (Sl.2) su:



Slika 2. Stvaranje prediktivnog modela iz podataka za obuku i algoritma

- Uzorak podataka: podaci koji se prikupljaju, koji opisuju određeni, ciljani problem s poznatim odnosima između ulaza i izlaza.
- Obuka modela: algoritam koji se koristi na podacima uzorka da bi se izradio model koji se kasnije može koristiti.
- Donošenje predikcije: Korištenje naučenog modela na nove podatke za koje nije poznat izlaz (Sl. 3).



Slika 3. Korištenje modela u cilju predviđanja o novim podacima

Cilj algoritama nadziranog učenja je da korištenjem podataka sa poznatim odnosima/vezama izradi model tih odnosa sa ciljem da predvidi, napravi predikciju. Ako je izlaz numerička vrijednost, onda to predstavlja problem regresije, a ukoliko je izlaz kategorija, kao što je u slučaju ovoga rada, onda to spada u vrstu problema klasifikacije, koja, uz regresiju, predstavlja važnu klasu problema prediktivnog modeliranja.

A. Binarna klasifikacija

Binarna klasifikacija u osnovi je zadatak razvrstavanja elemenata određenog skupa u dvije grupe (postoje dvije klase) na temelju prethodno određenih pravila klasifikacije, na osnovu toga da li neki predmet ima ili nema neke određene karakteristike na primjer [8]. Stvarni izlaz mnogih algoritama za binarnu klasifikaciju je predviđanje rezultata. Binarna klasifikacija u kontekstu ovoga rada je korištena da se predvidi vjerovatnoća da će oprema - avionski motor imati kvar u nekom budućem periodu (izraženom kroz broj ciklusa). Generalno posmatravši taj vremenski period je determinisan na poslovnim pravilima i podacima sa kojim se raspolaže. U praksi neki uobičajeni vremenski periodi su minimalno vrijeme realizacije (*lead time*) potrebno da se nabave rezervni dijelovi vjerovatno oštećene komponente ili vrijeme koje potrebno da se primjene postojeći resursi održavanja kako bi se izvršile rutine, uobičajeni postupci održavanja za slučaj rješavanja problema koji će se vjerovatno (po predikciji) desiti u okviru toga vremenskog perioda. Taj period za koji se vrši predviđanje uobičajeno se naziva horizont (vidokrug) period "X".

Da bi se vršila binarna klasifikacija, potrebno je da se identifikuju dva tipa primjera, koji se mogu nazvati pozitivan i negativan. U posmatranom kontekstu generalno uzevši, svaki primjer je zapis koji pripada vremenskoj jedinici za neku opremu, a koji konceptualno i sažeto opisuje njene operativne uslove do tog vremena, posredstvom inženjeringa karakteristika, koristeći istorijske i druge izvore podataka. U kontekstu binarne klasifikacije, pozitivni tipovi primjera obilježavaju, odnosno označavaju kvar (vrijednost oznake klase „1“), dok svaki negativan tip označava normalne operacije (vrijednost oznake „0“), pri čemu su oznake kategoričkog tipa. Cilj je u biti pronaći (najbolji) model koji identifikuje svaki novi primjer kao kvar ili kao normalnu operaciju u u roku od narednih X jedinica vremena.

Da bi se kreirao jedan takav prediktivni model, koji će u osnovi dati vjerovatnoću dešavanja kvara opreme u sledećih X vremenskih jedinica (gdje je „X“ broj perioda vremena koji je od interesa) potrebno je izvršiti i označavanje primjeraka, pri čemu je za potrebe binarne klasifikacije neophodno izvršiti označavanje zapisa oznakama koje su kategoričke varijable (npr. Klasa {0,1} u okviru primjera datog eksperimentalnom postavkom).

IV. EKSPERIMENTALNE POSTAVKE I REZULTATI

A. Priprema podataka za trening/ testiranje

Kao polazni skup podataka u ovom radu je korišten "Turbofan Engine Degradation Simulation Data Set" javno dostupan (doniran od NASA, C-MAPSS simulator sistema) set podataka koji predstavlja simulaciju degradacije avionskog motora. Za testiranje je korišten jedan od četiri dostupna simulirana različita seta (svaki je nadalje sadržan od tri (pod)seta), koji su izvorno dostupni kao zip-kompresovana tekstualna datoteka [9]. U osnovi u okviru rada biće korištena tri tipa seta podataka kao ulaz:

- Trening ili podaci za obučavanje modela. To su podaci avionskog motora radom do otkaza (u principu ovdje su podaci sve do momenta kada se desio kvar)
- Testni podaci ili podaci za testiranje – Ovo su operativni podaci koji u sebi ne sadrže zapis o kvaru. (Imaju istu strukturu, tj šemu podataka kao i trening podaci).
- Referentni (ground truth) podaci – Oni sadrže tačnu informaciju o preostalom periodu vremena/ciklusa za svaki motor koji se nalazi u testnim podacima.

U kontekstu prethodne podjele, za trening je korišten, Train_FD003.txt (radi lakše čitljivosti svi nazivi atributa instanci semantički su prilagođeni adekvatnim terminima našeg jezika; na Sl.1 se može vidjeti dio atributa (u nastavku su atributi Sensor3-Senzor21)). Ovi podaci [10] se sastoje od višestrukih multivarijantnih vremenskih serija, ciklusom kao vremenskom jedinicom zajedno sa očitanim vrijednostima 21 senzora za svaki ciklus, uz oznaku motora i tri različite postavke/setovanja za svaki motor. Svaka vremenska serija je generisana od strane različitih motora istog tipa (podaci se mogu smatrati da pripadaju floti motora iste vrste). Svaki motor započinje sa inicijalnom izdržljivošću i varijacijama u proizvodnji koji su nepoznati korisniku. U ovim simuliranim

podacima se smatra/podrazumjeva da će svaki motor na početku svake vremenske serije da radi normalno. On počinje sa degradacijom svojih karakteristika u nekoj tački tokom serije operativnih (radnih) ciklusa. Taj degradacioni proces raste u obimu tokom vremena. Kada se postigne prethodno definisani prag (treshold), tada se motor smatra nesigurnim za buduće operacije. Drugim riječima, zadnji ciklus u svakoj vremenskoj seriji se smatra tačkom kvara za određeni motor. Ako se pogledaju dostupni podaci može se zaključiti da je motor sa IDmotora=1 imao kvar u 259 ciklusu (Sl.1). Set podataka za trening, Train_FD003.txt ima 24720 instanci i 26 atributa (inicijalno podaci nisu označeni, ne postoji oznaka klase (class)).

Relation: Serije_podataka_za_trening							
1: ID_motora	2: Ciklus	3: Setovanje1	4: Setovanje2	...	25: Senzor20	26: Senzor21	27: Klasa
Numeric	Numeric	Numeric	Numeric	...	Numeric	Numeric	Nominal
1.0	246.0	5.0E-4	2.0E-4		38.4	23.1525	1
1.0	247.0	0.0012	2.0E-4		38.5	23.1023	1
1.0	248.0	-0.0043	-1.0E-4		38.41	23.1492	1
1.0	249.0	0.0017	4.0E-4		38.38	23.2023	1
1.0	250.0	0.0036	1.0E-4		38.69	23.1101	1
1.0	251.0	0.0021	0.0		38.46	23.0971	1
1.0	252.0	-5.0E-4	-3.0E-4		38.36	23.0905	1
1.0	253.0	0.0012	-4.0E-4		38.34	23.1021	1
1.0	254.0	-0.001	-4.0E-4		38.36	23.1083	1
1.0	255.0	0.005	-4.0E-4		38.39	23.0716	1
1.0	256.0	0.0019	1.0E-4		38.46	23.0488	1
1.0	257.0	-0.0032	-2.0E-4		38.36	23.1058	1
1.0	258.0	-5.0E-4	1.0E-4		38.49	23.0658	1
1.0	259.0	0.0043	-1.0E-4		38.35	23.0707	1
2.0	1.0	0.0034	3.0E-4		39.04	23.3588	0
2.0	2.0	0.0031	3.0E-4		38.96	23.4683	0
2.0	3.0	-7.0E-4	5.0E-4		38.93	23.3825	0
2.0	4.0	0.0049	-4.0E-4		38.85	23.3706	0
2.0	5.0	3.0E-4	-1.0E-4		39.23	23.4144	0
2.0	6.0	-0.0012	2.0E-4		39.13	23.4867	0
2.0	7.0	0.0051	-4.0E-4		39.05	23.3851	0
2.0	8.0	5.0E-4	1.0E-4		39.03	23.4174	0
2.0	9.0	-4.0E-4	4.0E-4		38.93	23.4609	0

Slika 4. Dio originalnog seta trening podataka prilagođen .arff formatu (naznačen zadnji ciklus 259 za IDmotora=1, naznačene oznake klase (nisu dio originalnog seta) naknadno kreirane za potrebe binarne klasifikacije)

Za testiranje je korišten Test_FD003.txt. Testni podaci imaju istu šemu podataka kao i trening podaci. Jedina razlika leži u tome da oni ne sadrže podataka kada se desio kvar. Drugim riječima zadnje vrijeme tj. zadnji ciklus ne predstavlja tačku kvara. Tako npr. u ovim testnim podacima motor sa IDmotora=1 je radio od ciklusa 1 do ciklusa 233 (Sl 5.) (to nije

Relation: Serije_testnih_podataka							
1: ID_motora	2: Ciklus	3: Setovanje1	4: Setovanje2	...	25: Senzor20	26: Senzor21	27: Klasa
Numeric	Numeric	Numeric	Numeric	...	Numeric	Numeric	Nominal
1.0	222.0	3.0E-4	1.0E-4		39.31	23.4482	0
1.0	223.0	-5.0E-4	4.0E-4		39.32	23.4728	0
1.0	224.0	-0.0025	2.0E-4		39.25	23.5433	0
1.0	225.0	0.003	-1.0E-4		39.31	23.6537	0
1.0	226.0	0.0016	-4.0E-4		39.25	23.6287	0
1.0	227.0	-5.0E-4	2.0E-4		39.28	23.5238	0
1.0	228.0	-8.0E-4	0.0		39.38	23.6203	0
1.0	229.0	0.003	1.0E-4		39.22	23.4479	0
1.0	230.0	0.0018	0.0		39.37	23.6093	0
1.0	231.0	-2.0E-4	4.0E-4		39.34	23.616	0
1.0	232.0	0.0038	-5.0E-4		39.21	23.5743	0
1.0	233.0	0.0021	2.0E-4		39.43	23.5679	0
2.0	1.0	-0.0017	-2.0E-4		39.01	23.4496	0
2.0	2.0	6.0E-4	2.0E-4		38.85	23.319	0
2.0	3.0	0.002	-1.0E-4		38.91	23.2714	0
2.0	4.0	-0.0019	-1.0E-4		39.01	23.3981	0
2.0	5.0	-0.0039	5.0E-4		38.98	23.4233	0
2.0	6.0	-0.0023	5.0E-4		39.09	23.3307	0
2.0	7.0	6.0E-4	-5.0E-4		39.01	23.2643	0
2.0	8.0	2.0E-4	-1.0E-4		38.87	23.3523	0
2.0	9.0	-0.0051	1.0E-4		38.84	23.3993	0
2.0	10.0	4.0E-4	0.0		38.79	23.4543	0

Slika 5. Dio originalnog seta testnih podataka u arff formatu (naznačen zadnji ciklus 233 za IDmotora=1, naznačene oznake klase (nisu dio originalnog seta), kreirane korištenjem referentnih podataka, za potrebe testiranja modela predikcije putem binarne klasifikacije)

zadnji ciklus, nije pokazano koliko je još ciklusa motor imao prije kvara). Dakle razlika između trening seta podataka i testnog seta podataka je samo u potpunosti ciklusa do pojave kvara. Set podataka za testiranje, Test_FD003.txt ima 16596 instanci i 26 atributa.

RUL_FD003.txt sadrži referentne podatke (jednodimenzionalna tabela, vektor). Ovi referentni podaci obezbjeđuju broj preostalih radnih ciklusa za motor koji se nalazi u testnim podacima. U ovom slučaju ovi podaci sadrže ukupno 100 vrijednosti (za svaki Idmotora; predstavljaju preostali broj ciklusa za svaki IDmotora koji se nalazi u testnim podacima). Tako npr. u prvom redu se nalazi vrijednost 44, što znači da je motor sa IDmotora=1 imao još 44 ciklusa dodatno na ona 233 ciklusa iz testnih podataka (Sl.5), prije kvara.

1) Inženjering karakteristika

Inženjering karakteristika u kontekstu predikcije stanja određene opreme (motora, mašine,...) polazi od ideje da se u osnovi konceptualno opiše i sažme stanje posmatrane opreme u datom vremenskom trenutku koristeći istorijske (prošle) podatke koji su prikupljeni do tog trenutka vremena [11]. U kontekstu takvog problema veoma su važne tzv. lag karakteristike (lag, usporavanje, kašnjenje, pomaknuto ..), koje se generišu na osnovu postojećih zapisa vremenskih serija originalnog seta podataka.

Karakteristike koje će biti uključene u trening podacima mogu biti grupisane u dvije kategorije. Selektovani su svi sirovi atributi, atributi koji su uključeni u originalne ulazne podatke. Pored njih kreirani su i korišteni i agregirani atributi, atributi koji u osnovi sažimaju istorijske aktivnosti za posmatranu problematiku vezanu za avionski motor. U okviru eksperimentalne postavke korišteni parametar vremenskog prozora W iznosi 7 (W=7) za koji se vrši agregiranje trening skupa. Konkretno, u okviru eksperimentalne postavke ovoga rada kreirana su setovi dva tipa agregiranih karakteristika za svaki od 21 senzora podataka:

- SSenzor1–SSenzor21: predstavljaju kretanje prosječne vrijednosti – srednja vrijednost sa senzora u najviše W prethodnih ciklusa (u odnosu na posmatrani),
- SDSenzor1–SDSenzor21: standardna devijacija senzorskih vrijednosti u najviše W prethodnih ciklusa,

a koji su uključeni u trening skup podataka. Posmatrano sumarno kroz broj atributa to znači dodatnih 42 atributa u odnosu na početni (originalni) koji ima 26 atributa, odnosno 69 atributa ukupno za pomenuta trening skup (uračunat i atribut klase koji će se naknadno kreirati). Analogan postupak se odvija i na testnim podacima.

2) Označavanje instanci

Predikcija putem binarne klasifikacije korištena u ovom radu, u kontekstu prirode izabranog seta podataka, se odnosi na predviđanje kvara avionskog motora u opsegu određenog broja ciklusa, konkretnije predviđanje kvara je rađeno za opseg od 30 ciklusa (ciklus je u tom smislu mjerna jedinica vremena). Neophodan korak (već je pomenuto da trening skup (niti testni analogno tome) izvorno ne sadrži oznaku klase) da bi se izvršila predikcija putem binarne klasifikacije je označavanje klase[12]. Oznaka klase za potrebe binarne klasifikacije

(ilustracija za trening set na Sl. 4) je urađeno uzimajući 30 zapisa (instanci, primjeraka) i označavajući ih sa „blizina kvara“ - vrijednost oznake je 1 (Sl. 4), dok je označavanje za sve ostale zapise „normalno“ tj. nema kvara, vrijednost oznake je 0 (prethodno kreiran atribut Klasa {0,1}). Označavanje testnih podataka Test_FD003.txt (Sl. 5) je izvršeno na bazi referentnih podataka RUL_FD003.txt, pri čemu, po analogiji, prolaze kroz istu proceduru označavanja kao i označavanje trening podataka.

Za potrebe kreiranja finalnih verzija trening i testnog seta podataka i procesa označavanja klase u skladu sa odabranom vrijednošću predviđanja, razvijena je adekvatna programska podrška (C#), a tako kreirani trening i testni skupovi korišteni su u okviru Weka alata (.arff formalizovan Weka format) za data mining - korištenjem izvedbe Weka alata prethodno pomenutih/odabranih klasifikatora.

B. Korišteni klasifikatori

Za potrebe izgradnje prediktivnog modela, u okviru Weka aplikacije, korišteno je šest različitih klasifikatora: algoritam logističke regresije (*Logistic Regression*), algoritam temeljen na Bayesovim mrežama Naivni Bajesov (*Naive Bayes*), algoritam stabla odlučivanja slučajna šuma (*Random Forest*), algoritma k-najbližeg susjeda IBk, implementacija algoritma vektora podrške (SMO) i višeslojni perceptron (MLP, Multilayer Perceptron). Logistička regresija (u osnovi koje leži logistička funkcija – poznata i kao sigmoid funkcija) svakako je jedan od najpopularnijih algoritama mašinskog učenja za binarne klasifikacije, jer je to u biti jednostavan algoritam koji postiže dobre rezultate na širokom rasponu problema. Naivni Bajesov algoritam je jedan od najpoznatijih modela koji su temeljeni na osnovu Bajesovih mreža (grafičke strukture za predstavljanje uslovnih vjerovatnoća između velikog broja varijabli/atributa te donošenje uslovljenih zaključaka vezano za navedene varijable/atribute). Glavna osobina algoritma slučajnih šuma je da on pri raspoznavanju ne koristi samo jedan klasifikator nego kreira veći broj klasifikatora u obliku stabla odlučivanja (*decision trees*) pri čemu svako od njih učestvuje u formiranju konačnog rezultata.

U eksperimentu je korišten i IBk klasifikator koji predstavlja algoritam k najbližeg susjeda (primjer lijernih (*lazy*) klasifikacijsko-predikcijskih postupaka). Prosti algoritam najbližeg susjeda IB1 koristi normalizovano Euklidsko rastojanje (najkraća udaljenost između dvije tačke u jednom prostoru), dok algoritam k najbližeg susjeda IBk može selektovati odgovarajuću vrijednost za k baziranu na kros

validaciji (u korištenoj implementaciji k=7). Korišten je i SMO algoritam koji predstavlja (Weka) implementaciju vektora podrške SVM (*support vector machines*), implementacija koja globalno zamjenjuje sve vrijednosti koje nedostaju i pretvara nominalne attribute u binarne (primjenjuje Džon Platt (John Platt) algoritam sekvencijalne minimalne optimizacije).

Višeslojni perceptroni predstavljaju najistaknutiji tip vještačke neuronske mreže i pripadaju klasi mreža sa prostiranjem unaprijed, tzv. *feedforward* mreže, koje ne sadrže nikakve cikluse (opozit su periodične neuronske mreže koje imaju cikluse). U radu će biti korištena implementacija MLP Weka alata Ova funkcija implementira algoritam povratnog prostiranja (*backpropagation*) za izgradnju modela neuronske mreže za klasifikaciju instance [13].

C. Evaluacija performansi izgrađenog modela

Procjena performansi kreiranih prediktivnih modela u okviru ove postavke izvršena je opcijom testiranja korištenjem svih dostupnih instanci testnog seta podataka (prethodno pripremljen u skladu sa pripremama izvršenim na trening podacima (neophodnost iste šeme)).

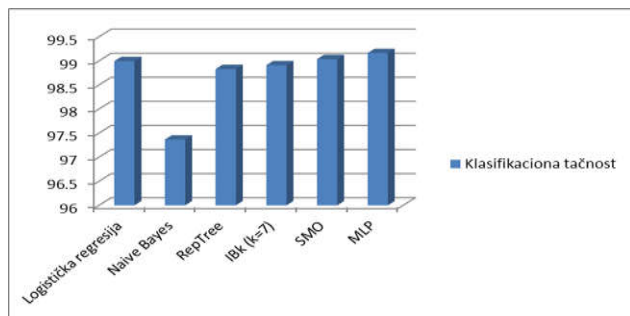
Fitness dobijenih modela, kao što se može vidjeti u Tabeli I, dat je sledećom (Weka dostupnom) metrikom: klasifikaciona tačnost (*classification accuracy*) izražena u procentima (ispod je radi ilustracije, u zagradama, pridodat broj pogrešno klasifikovanih instanci od ukupno 16596 koje su korištene za testiranje); koeficijent Kappa statistike (*Kappa statistics*) kao mjera slaganja predikcije sa pravom klasom (1.0 označava potpuno slaganje); srednja apsolutna greška MAE (*mean absolute error*); korijen srednje kvadratne greške RMSE (*root mean square error*); relativna apsolutna greška RAE (*relative absolute error*); korijen relativne kvadratne greške RRSE (*root relative squared error*). (Niže vrijednosti grešaka znače da je model tačniji u donošenju predviđanja).

D. Postignuti rezultati i diskusija

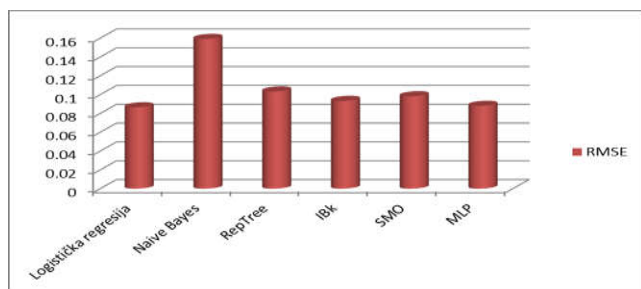
U Tabeli I, korištenjem pomenute metričke prikazani su postignuti rezultati prediktivnih modela svih šest korištenih klasifikatora: logistička regresija, Naive Bayes, Random Forest, IBk, SMO i MLP respektivno, dobijeni treniranjem nad kompletnim setovima za trening i evaluacijom korištenjem kompletnog seta podataka za testiranje, pri čemu su u testnim podacima korištene sve vremenske serije.

TABELA I. EKSPERIMENTALNI REZULTATI PREDIKTIVNIH MODELA KORIŠTENJEM MODELA BINARNE KLASIFIKACIJE (ŠEST RAZLIČITIH ALGORITAMA MAŠINSKOG UČENJA)

	Algoritmi mašinskog učenja					
	<i>Logistička regresija</i>	<i>Naive Bayes</i>	<i>Random Forest</i>	<i>IBk (k=7)</i>	<i>SMO</i>	<i>MLP</i>
Klasifikaciona tačnost (%) (broj netačno klasificiranih)	98.9937 (167)	97.3668 (437)	98.9998 (166)	98.9094 (181)	99.0359 (160)	99.1624 (139)
Kappa koeficijent	0.6706	0.4998	0.6356	0.6166	0.6764	0.7269
MAE	0.012	0.0264	0.0136	0.0141	0.0096	0.0084
RMSE	0.0864	0.1591	0.0857	0.0933	0.0982	0.088
RAE (%)	9.0544	19.906	10.2439	10.6032	7.2649	6.3668
RRSE (%)	53.558	98.5757	53.0751	57.8347	60.834	54.5353



Slika 6. Grafički prikaz klasifikacijske tačnosti prediktivnih modela korištenih klasifikatora



Slika 7. Grafički prikaz vrijednosti korijena srednje kvadratne greške prediktivnih modela korištenih klasifikatora

Na osnovu obavljenih eksperimenata, uvidom u dobijene i prikazane rezultate date Tabelom I, može se zaključiti da su se izdvojila dva klasifikatora. U pitanju su algoritam logističke regresije i višeslojni perceptron koji su na određeni način „podjelili“ korištenu metriku u svoju korist.

Posmatrano u kontekstu klasifikacione tačnosti (Sl. 6), vrijednosti Kappa koeficijenta, te vrijednosti srednje apsolutne greške MAE i relativne apsolutne greške RAE, najbolje rezultate (99.1624%, 0.7269, 0.0084 i 6.3668%, respektivno) postigao je predstavnik neuronskih mreža višeslojni perceptron klasifikator. Klasifikator logističke regresije postigao je bolji rezultat za vrijednost, prikazane na Sl. 6, korijena srednje kvadratne greške RMSE (sa 0.0864 neznatno bolji od 0.088 postignuto od strane klasifikatora višeslojni perceptron) i korijena relativne kvadratne greške RRSE u iznosu od 53.558%. Generalno, može se reći da je najbolji prediktivni model održavanja za postavljeni problem predikcije, u okviru korištenih postavki, postigao model neuronske mreže - algoritam višeslojni perceptron (najlošije rezultate ostvario je Naivni Bajesov klasifikator).

V. ZAKLJUČAK

U radu su, uz neophodan teorijski okvir za potpunije sagledavanje i razumjevanje primjenjenog pristupa i postupaka, korišteni i poređeni različiti algoritmi mašinskog učenja u okviru datih eksperimentalnih postavki, sa ciljem da bi se kreirao model prediktivnog održavanja, koji u osnovi daje vjerovatnoću dešavanja kvara opreme za neki sledeći broj vremenskih jedinica - predviđanje kvara avionskog motora u opsegu određenog broja ciklusa. Dalja istraživanja i eksperimentisanja (apriori odgovor ne postoji) u kontekstu prezentovanog pristupa moguća su u domenu inženjeringa

karakteristika (u rada su korištena dva tipa agregiranih karakteristika) i selekcije atributa na primjer.

LITERATURA

- [1] J. Levitt, Complete Guide to Predictive and Preventive Maintenance, Industrial Press, New York, 2003.
- [2] O. Janković, “M2M komunikacija u službi prediktivnog održavanja“, 2nd International Scientific Conference COMETA 2014, “COMETA 2014 - Conference on Mechanical Engineering Technologies and Application”, Zbornik Radova, str.725-728, Jahorina, 2014.
- [3] P. Bastos, I. Lopes, and L. Pires, A Maintenance Prediction System using Data Mining Techniques, Proceedings of the World Congress on Engineering 2012 Vol III, WCE 2012, July 4 - 6, 2012, London, U.K.
- [4] P. Bastos, I. Lopes, L. Pires, Application of data mining in a maintenance system for failure prediction, Safety, Reliability and Risk Analysis: Beyond the Horizon – Steenbergen et al. (Eds), 2014.
- [5] W. Sammouri. Data mining of temporal sequences for the prediction of infrequent failure events: application on floating train data for predictive maintenance. Signal and Image processing. Universit'e Paris-Est, 2014.
- [6] J.M. Gross, Fundamentals of Preventive Maintenance, AMACOM, New York, 2002.
- [7] R. K. Mobley, An Introduction to Predictive Maintenance, Elsevier Science USA, 2002.
- [8] I. H. Witten, and E. Frank, Data Mining Practical Machine Learning Tools and Techniques, Elsevier, 2011.
- [9] A. Saxena, and K. Goebel, "Turbofan Engine Degradation Simulation Data Set", NASA Ames Prognostics Data Repository, (<http://ti.arc.nasa.gov/project/prognostic-data-repository>), NASA Ames Research Center, Moffett Field, CA, 2008.
- [10] A. Saxena, K. Goebel, D. Simon, and N. Eklund, —Damage propagation modeling for aircraft engine run-tofailure simulation—. In International Conference on Prognostics and Health Management, pages 1–9, 2008.
- [11] O. Janković, “Inženjering karakteristika u kontekstu predikcije korištenjem regresije”, INFOTEH 2017, podnesen za objavljivanje
- [12] F. U. Boylu, “ Predictive Maintenance Modelling Guide R Notebook, Cortana Intelligence Gallery“, <https://gallery.cortanaintelligence.com/Notebook/Predictive-Maintenance-Modelling-Guide-R-Notebook1.2016>.
- [13] O. Janković, “Primjena i evaluacija klasifikatora višeslojnog perceptrona za potrebe klasifikacije linearno neseparabilnih problema“, SYM-OP-ISA 2015, str. 208-211, Ivanjica, 2015.

ABSTRACT

The focus of this work is the implementation process of predictive maintenance model, considering the nature of the observed problems - the prediction on the existence of failure of an aircraft engine within a certain number of operating cycles, in the context of binary classification model. To create and at the end compare the achieved predictive models, the necessary training and test data sets on the basis of a simulation of the degradation of an aircraft engine data set are created, using different types of machine learning algorithms (using logistic regression, the algorithm based on Bayesian networks, decision tree algorithm, k-nearest neighbors algorithm, the implementation of support vector algorithm and multilayer perceptron).

IMPLEMENTATION OF THE PREDICTIVE MAINTENANCE MODEL USING BINARY CLASSIFICATION

Olivera Janković