

Doprinos pouzdanosti predikcije za prediktivni model održavanja avionskog motora

Olivera Janković
ORAO a.d.
Bijeljina, RS, BiH
olivera.jankovic@orao.aero

Đorđe Babić
RAF
Beograd, Srbija
djbabic@raf.rs

Sažetak—Bez obzira na primjenjene tehnike mašinskog učenja, prilikom izgradnje prediktivnih modela temeljenih na podacima, imperativ svakog modelara je pouzdana predikcija. Pouzdanost predikcije ogleda se u osnovi kroz mogućnost prediktivnog modela da performanse predikcije, postignute na podacima na kojima je model obučavan, potvrdi i na prethodno nepoznatim testnim podacima, sa ciljem, da se u konačnici iste reprodukuju i na podacima koji će se koristiti nakon aplikacije modela. U radu će biti prikazana primjena pristupa na kojem se temelji povjerenje doprinosa pouzdanosti prediktivnog modela, implementirana za potrebe pouzdanosti predikcije prediktivnog modela održavanja avionskog motora, posrednim doprinosom pouzdanosti korištenih klasifikatora binarne klasifikacije.

Gljučne riječi—mašinsko učenje, binarna klasifikacija, predikcija, prediktivni model održavanja avionskog motora;

I. UVOD

Sudionici smo vremena napretka i sveprisutne dostupnosti računarskog hardvera i posljedično tome i raspoloživosti velikih količina podataka različitih izvora, što udruženo sa razvojem i dostupnošću raznih algoritama i alata mašinskog učenja otvara mogućnosti za kreiranje prediktivnih modela baziranih na podacima (*data driven models*).

Cijela priča, dakle, počinje sa podacima i potrebom da se isti iskoriste za potrebe predviđanja, nakon što se u njima otkriju određene zakonitosti i strukture. Uspješnost procesa pronalazjenja zakonitosti i struktura, skrivenih obrazaca (*patterns*) opet je ovisna o karakteristikama podataka, kojima raspolazu oni, koji za cilj imaju da kreiraju prediktivni model. Na tom putu, pored ostalog, je i odabir adekvatnih klasifikacionih ili regresionih tehnika mašinskog učenja, kao i nesporan značaj strategije evaluacije, bazirane na adekvatnoj, objektivnoj metrici na primjer. U biti, radi se o mnogo međusobno ovisnih faktora koji za potrebe postizanja dobrih rezultata traže posvećenost detaljnom uvidu razumjevanja svega pomenutog, uz mnogo pratećeg eksperimentalnog rada i mnoštvo eksperimentalnih postavki.

Dostupnost modernih računarskih alata za potrebe istraživanja podataka dovela je, sa jedne strane do masovnosti primjene samog procesa, dok su se na drugoj strani našli upitnosti kredibiliteta i pouzdanost modela. Naime, obzirom na relativno pojednostavljen niz koraka (izbor tehnike

modelovanja, odabir i uključenje raspoloživih podataka i generisanje predikcije) koji može dovesti do kreiranja prediktivnih modela, i bez prethodnog poznavanja konteksta i domena problema i razumjevanja uticaja prediktora na model, bez postupka preprocesiranja, bez razumjevanja distribucije ishoda, ..., jednom riječju do prediktivnih modela, najčešće upitne vjerodostojnosti i pouzdanosti, u kontekstu željenog cilja, potrebe stabilne i pouzdane predikcije novih uzoraka.

Primjena tehnika mašinskog učenja u oblasti prediktivnog održavanja u opštem, širem smislu, pa shodno tome i avionskog motora u užem smislu predstavlja aktuelno i nedovoljno istraženo područje. U prilog tome ide i dato pregledno stanje u radu [1] u kome se vrši istraživanje na više od šezdeset radova objavljenih tokom prethodećih mu pet godina, fokusirajući se na pristupe za rješavanje problema vezanih za otkrivanje kvarova i probleme prediktivnog održavanja korištenjem pristupa mašinskog učenja. Naime, po zapažanju samog autora, iznenađujuće je da, ključni elementi postupka inženjeringa karakteristika (*engineering characteristics*): područja selekcije karakteristika (*feature selection*) i ekstrakcije karakteristika (*feature extraction*), te područje označavanje podataka (*data labeling*) nisu dovoljno dobro istraženi u korištenim radovima; pri čemu stoji stav autora da su ovi koraci veoma važni za odabir adekvatnog pristupa mašinskog učenja sa analitičke tačke gledišta i da utiču na dobru tačnost sistema.

Problemi sa kojim se treba nositi prilikom izgradnje prediktivnih modela je i prevelika prilagođenost modela (*overfitting*), koju treba izbjeći obzirom da usled koje dolazi do upitne pouzdanosti. Uzimajući u obzir sve pomenuto, u cilju podizanja povjerenja pouzdanosti, u okviru dijela eksperimentalne postavke biće prikazan, na način prilagođen tom cilju, postupak podešavanja parametara klasifikatora metode vektora podrške i algoritma slučajna šuma, za potrebe izgradnje, a u krajnjem odabira što tačnijeg ali i što pouzdanijeg, prediktivnog modela održavanja avionskog motora.

II. PREDIKTIVNI MODELI ODRŽAVANJA

Pristupi mašinskog učenja (*Machine learning approaches*), odnosno mašinsko učenje [2], pored ostalog omogućavaju analizu prikupljenih podataka uz smanjenje troškova

računanja, obzirom da sama obimnost podataka utiče na nepraktičnost manuelne analize. Jedan od primjera korištenja algoritama mašinskog učenja, analizirano iz perspektive scenarija prediktivnog održavanja, dat je u [3] (analize primjene su korisne, obzirom da trenutno nije dostupan pristup za donošenje odluke za odabir tehnika mašinskog učenja).

A. Zahtjevi na ulazne podatke

Ono što karakteriše metode temeljene na podacima (*data-driven*) [4] je da pored kvaliteta i veličine uzorka podataka, zahtijevaju dovoljno primjeraka koji su se izvršavali do kvara (*run-to-failure*), odnosno koriste obrasce degradacije dovoljnog broja uzoraka koji predstavljaju progresiju greške opreme. Ovaj zahtjev predstavlja najveći izazov u prognozama temeljenim na podacima, jer često, iz objektivnih razloga, nije moguće dobiti uzorke progresija kvara. Industrijski sistemi tako na primjer ne smiju raditi do kvara zbog posljedica koje mogu zadobiti, a zahtjev sigurnosti je tu kao neupitan razlog kada je riječ o avioindustriji. Naime, cilj je osigurati pravilan rad avionskog motora, u svim uslovima, sa nultom vjerovatnoćom neuspjeha [5]. Za potrebe primjene mašinskog učenja, stoga se najčešće koriste dostupni setovi podataka iz javno dostupnih repozitorijuma podataka (najpoznatiji je svakako UCI, [6]). U okviru NASA Ames repozitorija podataka [7] za predikciju (2017.) je rastući izvor koji obuhvaća nekoliko skupova prognostičkih podataka koje doniraju univerziteti, kompanije ili agencije. Skupovi podataka u repozitoriju se sastoje od podataka vremenskih serija koji prikazuju slučajeve radom do otkaza. Na raspolaganju su kompleti skupova podataka prognostičkih podataka. Jedan od takvih skupova koji se planira koristiti kao polazna osnova za ulazni skup predviđene eksperimentalne postavke za izgradnju prediktivnog modela održavanja avionskog motora, odnosno povećanju pouzdanosti predikcije istog, je „Turbofan Engine Degradation Simulation Dataset“. Sam postupak simulacije, od strane vlasnika skupova, prikazan je u [8].

U radu [9] koji se u osnovi bavi benčmarkom određenog broja javno dostupnih setova podataka za njihovu podobnost u testiranju prognostičkih pristupa, očekuje se da skup podataka koji će se koristiti u svrhu predikcije mora imati minimalnu veličinu uzorka od oko 10, kako bi se efikasno primjenilo modeliranje bazirano na podacima, pri čemu modeli temeljeni na podacima zahtijevaju statistički dovoljan broj tzv. kvarom do otkaza uzoraka. Autor analizira i set podataka „Turbofan Engine Degradation Simulation Dataset“, pomenutog NASA Ames repozitorija. Prema [9] model koji su vlasnici skupova podataka primjenjivali je eksponencijalna degradacija prikazana u jednačini (1) gdje je (d) inicijalna degradacija, (A) faktor skaliranja, (B(t)) vremenski varijabilni eksponent, i (th_w) gornji prag habanja. Model je opšta jednačina uobičajenih modela širenja štete (npr. Arrhenius model).

$$h(t)=1-d-A\exp^{B(t)/th_w} \quad (1)$$

Stav autora, koji je radio benčmark, je da skup podataka ispunjava uvjete za pristup koji se temelji na podacima, budući da su s podacima dostupni dovoljni podaci i vrijednosti za preostali korisni vijek upotrebe (Remaining Useful Life, RUL), vrijednosti koje se procjenjuju u vremenskim jedinicama (npr. sati ili ciklusi).

B. Istraživanja primjene vezano za avionske motore

U sledećim radovima naglasak je na raznovrsnim aplikacijama primjene, različitim načinima i pristupima u oblasti predikcije vezano za avionske motore. U radu [10] je razmatran problem vremenske predikcije kvara avionskog motora kao generičkog prognostičkog problema, pri čemu se nisu u većoj mjeri bavili problemom određivanja optimalnih karakteristika. Autor [11] je koristio samorganizujuću (Self-Organizing Maps, SOM) vrstu moćne neuronske mreža za analizu i vizuelizaciju visoko dimenzionalnih podataka, da omogući ekspertima pomoćni alat za laku vizuelizaciju evolucije podataka mjerenih na motorima. Evolucija je karakterisana sa trajektorijom u dvodimenzionalnoj SOM mreži, pri čemu su se abnormalnosti i kvarovi ogledali u devijacijama ovih trajektorija u odnosu na normalna stanja. Primjer korištenja neuronske mreže za obučavanje i testiranje dat je u [12], pri čemu su podaci preprocesirani u dva različita koraka, što se ogleda kroz uklanjanje podataka sa nedostajućim vrijednostima iz seta podataka, i korištenje normalne distribucije i uklanjanjem ekstremnih autlajera, čime se smanjila varijansa kompletnog seta podataka. U radu je prikazano korištenje Echo State Network (ESN), relativno novog tipa rekurentnih neuronskih mreža (Recurrent Neural Network (RNN)). Kao jedna od glavnih prednosti ESN je procedura je obučavanja koja je bazirana na jednostavnoj linearnoj regresiji za tačnu predikciju preostalog životnog vijeka. Predlog modela degradacije baziran na kernel analizi glavne komponente (KPCA) za pouzdanost turbofenskog motora dat je u [13] kao i procjena preostalog životnog vijeka motora. U [14] korišten je metod temeljen na podacima za RUL predikciju avionskog motora pod nepoznatim početnim habanjem. U [15] se razmatra problem izgradnje statističkih modela sajber fizičkih sistema koji koriste operativne podatke koristeći studiju slučaja motora. Ovi modeli služe kao dodatak modelima zasnovanim na fizici, koji možda neprecizno odražavaju operativne performanse sistema. Rad [16] je značajan, u kontekstu ovog rada, obzirom da se koriste binarni klasifikatori za prognozu zdravlja mlaznih motora, pri čemu koristi podatke Rolls-Royce kompanije.

Činjenica važna pominjanja, je da je vrijednost prediktivnog održavanja prepoznata i od strane velikih igrača avio industrije, kao što su na primjer najstarija operativna aviokompanija KLM [17], i Pratt & Whitney, svjetski poznat američki proizvođač aviona [18], a aktuelnost se ogleda i kroz članke na portalima koji su posvećeni avio industriji [19], [20].

III. PODEŠAVANJE PARAMETARA MODELA

Obzirom da se pronalaženje obrazaca, skrivenih struktura u postojećim podacima, postiže određenim tehnikama koje koriste vlastite parametre, koji im opet omogućavaju neophodnu fleksibilnost, u osnovi je potrebno pronaći najbolja setovanja parametara korištenih tehnika za raspoložive podatke. Opšte smjernice za pronalaženje parametara modela [21], koje se baziraju na izgradnji modela za različite vrijednosti parametara biće, za potrebe povećanja pouzdanosti prediktivnog modela održavanja avionskog motora, posredno preko korištenih klasifikatora, implementirane u narednom poglavlju. (Alternativa je korištenje algoritama, npr. simplex

metoda, koji imaju razvijene vlastite mogućnosti pronalaženja optimalnih parametara podešavanja).

Ključni momenat ovog pristupa je i izvršiti adekvatnu procjenu postignutih rezultata, i u tu svrhu se mogu koristiti neke od tehnika ponovnog uzorkovanja (*resampling*) kao što su: ponovljena k-unakrsna validacija, LOOCV, bootstrap i MonteKarlo; pri čemu se za odabir konačnih parametara modela u kontekstu odabira jednostavnijih modela može koristiti metoda jedne standardne greške ili izbor unutar određene tolerancije.

IV. EKSPERIMENTALNE POSTAVKE I REZULTATI

A. Set podataka prediktivnog modela za održavanje avionskih motora

```
% Vremenske serije turbo gasnih motora aviona
% trening skup, 24720 ins., bin.klasifikacija
% vremenski prozor agregiranja W=7

@relation Serije
@attribute ID motora numeric /Oznaka motora
@attribute Ciklus numeric /Vremenska jedinica
@attribute Setovanje1 numeric /Tri postavke,
@attribute Setovanje2 numeric /setovanja za
@attribute Setovanje3 numeric /svaki motor
@attribute Senzor1 numeric
@attribute Senzor2 numeric
@attribute Senzor3 numeric
@attribute Senzor4 numeric
@attribute Senzor5 numeric
...
@attribute Senzor16 numeric
@attribute Senzor17 numeric
@attribute Senzor18 numeric
@attribute Senzor19 numeric
@attribute Senzor20 numeric
@attribute Senzor21 numeric
...
@attribute SSenzor1 numeric
@attribute SSenzor2 numeric
@attribute SSenzor3 numeric
@attribute SSenzor4 numeric
@attribute SSenzor5 numeric
...
@attribute SSenzor16 numeric
@attribute SSenzor17 numeric
@attribute SSenzor18 numeric
@attribute SSenzor19 numeric
@attribute SSenzor20 numeric
@attribute SSenzor21 numeric
...
@attribute SSDenzor1 numeric
@attribute SSDenzor2 numeric
@attribute SSDenzor3 numeric
@attribute SSDenzor4 numeric
@attribute SSDenzor5 numeric
...
@attribute SSDenzor16 numeric
@attribute SSDenzor17 numeric
@attribute SSDenzor18 numeric
@attribute SSDenzor19 numeric
@attribute SSDenzor20 numeric
@attribute SSDenzor21 numeric
@attribute Klasa {0,1} Oznaka klase
```

Slika 1. Prikaz baznih i agregiranih karakteristika ulaznog seta podataka, korištenog u okviru eksperimentalne postavke, za potrebe pronalaženja optimalnih parametara prediktivnog modela

Za potrebe okvira ovoga rada biće korišten ulazni skup podataka čije su karakteristike prikazane na Sl.1. (karakteristike sirovih podataka su preuzete iz Train_FD003.txt odnosno Test_FD003.txt), a ostale su nastale procesom inženjeringa karakteristika, sa idejom da se poveća snaga predikcije u kontekstu krajnjeg cilja predviđanja [22]. Kako se radi o problemu binarne klasifikacije [23], kreirana je neophodna oznaka/labela klase, a samo označavanje je usko

vezano za period za koji se vrši predviđanje, tj. za 30 vremenskih jedinica/ciklusa. U skladu sa tim, 30 poslednjih instanci je označene sa 1, dok su sve ostale označene sa nula, što u suštini, respektivno, polazi od i ukazuje na blizinu i odsustvo kvara.

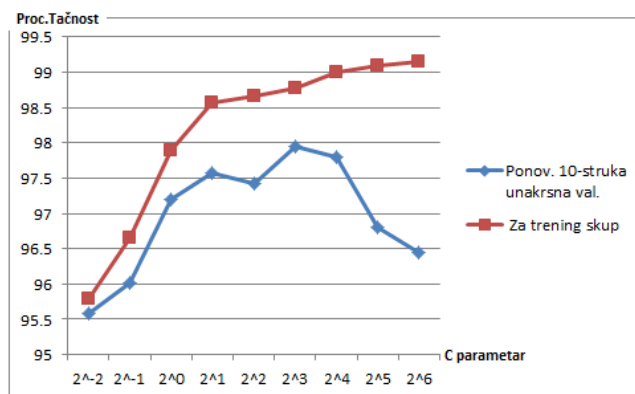
B. Proces podešavanje modela

Proces podešavanja parametara modela, procjenom koja je temeljena na tehnikama ponovnog uzorkovanja, korišten u okviru ove eksperimentalne postavke za potrebe povećanja pouzdanosti (nekih od) modela kandidata za izgradnju prediktivnog modela održavanja avionskog motora, sastoji se od sledećih koraka (korišteno je Weka Explorer GUI okruženje, a za potrebe pregleda tačnosti po fold-ovima izvršene su određene modifikacije koda korištenjem Java API (Sl.2)):

- Za svaki od korištenih modela odabran je inicijalni set vrijednosti parametara podešavanja:
 - Za izgradnju modela korištenjem SMO klasifikatora (implementacija vektora podrške u Weka), predviđeni set vrijednosti ($2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6$) za izabrani parametar cost (parametar c u Weka).
 - Za izgradnju modela korištenjem RandomForest klasifikatora (implementacija algoritma slučajna šuma u Weka) set poprima vrijednosti (1, 100, 200, 300, 500, 700, 900, 1100, 1300, 1500) za odabrani parametar broj stabala (parametar numIterations u Weka).
- Za svaki od navedenih klasifikatora, SVM i RandomForest, i vrijednost kandidata iz dodjeljenog mu seta:
 1. izvršiti ponovljenu, višestruku podjelu podataka; koristiti 5 puta ponovljenu 10-struku unakrsne validaciju (jedan od 10 dijelova iz podjele se ostavlja za procjenu performansi modela, evaluacija iz tačke 3);
 2. Izgradnja prediktivnog modela korištenjem SMO, odnosno RandomForest klasifikatora, za svaku od kandidat vrijednosti parametara c i numIterations respektivno;
 3. Evaluacija SMO/RandomForest prediktivnog modela, predikcija na izdvojenom setu (tačka 1); postupak procjene se ponavlja 50 puta, rezultati za svaki fold prikazani na Sl.2.
- Za svaki model pojedinačno izvršiti agregaciju postignutih vrijednosti predikcije, odnosno kreirati pripadajuće profile za oba, SVM i RandomForest model.
- Na osnovu dostupnih rezultata srednje tačnosti i standardne devijacije izračunati standardnu grešku i toleranciju (Tabela 1.), koje će se koristiti pri odabiru finalnih vrijednosti parametara podešavanja; parametar c (SMO) i numIterations (RandomForest).
- Izgraditi SMO i RandomForest prediktivne modele korištenjem odabranog konačnog parametra podešavanja, na kompletnom trening setu podataka.

Iteration	Fold	Percent correct
1	1	95.82
1	2	95.65
1	3	95.65
1	4	95.69
1	5	95.29
1	6	95.01
1	7	95.53
1	8	95.82
1	9	95.29
1	10	95.17
2	1	95.73
2	2	95.65
2	3	95.17
2	4	95.78
2	5	95.53
2	6	95.73
2	7	95.25
2	8	95.61
2	9	95.82
2	10	95.41
3	1	95.49
3	2	95.61
3	3	95.90
3	4	95.29
3	5	95.90
3	6	95.78
3	7	95.53
3	8	95.41
3	9	95.61
3	10	95.93
4	1	95.93
4	2	95.37
4	3	95.25
4	4	95.82
4	5	95.69
4	6	95.82
4	7	95.41
4	8	96.26
4	9	95.05
4	10	95.61
5	1	95.45
5	2	95.61
5	3	95.25
5	4	95.93
5	5	95.53
5	6	95.69
5	7	95.69
5	8	95.29
5	9	95.94
5	10	95.82

Slika 2. Prikaz rezultata za svaki fold pojedinačno, pedeset rezultata, 5 puta ponovljene 10-struke unakrsne validacije, za vrijednost parametra $cost=2^{-2}$ (prosječna tačnost = 95.59 i standardna devijacija=0.27)



Slika 3. Profil rezultata SMO modela; postignuto primjenom tehnike ponovnog uzorkovanja - 5 puta ponovljene 10-struke unakrsne validacije, i rezultati (optimistični) nastali evaluacijom korištenjem trening seta

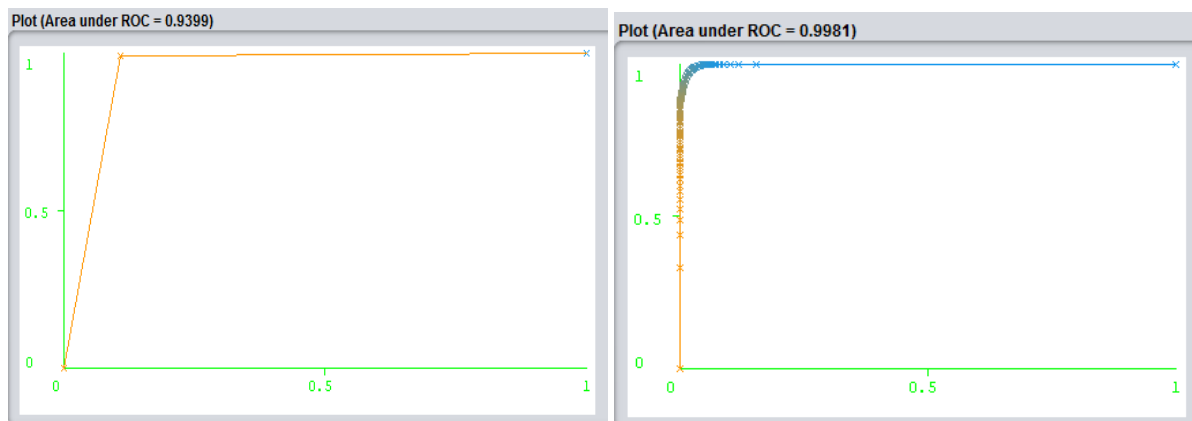
C. Izbor finalnih parametara

Nakon što je završen prethodno prikazani proces, određene performanse za svaku od vrijednosti kandidata - parametra c i $numIterations$ - potrebno je izvršiti odabir za krajnje setovanje. U tu svrhu kreirana je Tabela 1, sa elementima neophodnim za predložene pristupe odabira finalnih parametara.

Jedna od solucija, često najviše korištena, odabira finalnih parametara bila bi odabir vrijednosti parametra podešavanja za koju je prediktivni model postigao najveća tačnost. U ovom slučaju, kao što se može vidjeti (Tabela 1), brojčano najbolja performansa korištenog SVM modela, procjenjena tačnost od 97.95% postignuta je za vrijednost parametra $c=8$, dok je za RandomForest model tačnost od 98.85 postignuta za broj stabala 1300. Primjenom ovog pristupa na primjer, finalni prediktivni SMO model bi bio izgrađen sa vrijednošću 8 za parametar c i vrijednošću 1300 za $numIterations$ za RandomForest model.

TABELA I. REZULTATI SMO I RANDOMFOREST MODELA ZA RAZLIČITE VRIJEDNOSTI PRIPADAJUĆIH PARAMETARA, KORIŠTENJEM TEHNIKE PONOVRNOG UZORKOVANJA, 5 PUTA PONOVLJENE 10-STRUKE UNAKRSNE VALIDACIJE

Metoda vektora podrške (SMO)				Metoda slučajna šuma (Random Forest)			
Parametar c	Srednja vrijednost %	Std. greška %	Tolerancija %	Parametar $numIterations$	Srednja vrijednost %	Std. greška %	Tolerancija %
0,25	95,58	0,04	-2,41	1	96.63	0.27	-2.24
0,5	96,02	0,11	-1,97	200	98.10	0.33	-0.45
1	97,2	0,31	-0,76	300	98.23	0.35	-0.35
2	97,58	0,42	-0,38	500	98.44	0.38	-0.22
4	97,43	0,43	-0,53	700	98.50	0.37	-0.17
8	97,95	0,45	0	900	98.72	0.33	-0.13
16	97,8	0,43	-0,15	1100	98.80	0.34	-0.05
32	96,8	0,44	-1,17	1300	98.85	0.32	0
64	96,45	0,46	-1,63	1500	98.82	0.38	-0.03



Slika 4. Površina ispod ROC krive za SMO model (parametar $c=2$) i za RandomForest model (numIterations=900), ukazuje na izvanredno razdvajanje za oba modela (obje ROC krive sa slike se odnose na klasu 1)

Drugi pristup, poznat kao jedna standardna greška polazi od predhodno pronađene najbolje vrijednosti i pripadajuće joj jedne standardne greške. (Uvidom u profil tačnosti može se vidjeti da su to vrijednosti 97.95 i 0.45, odnosno 98.85 i 0.32 – zavisno od korištenog modela) Nadalje se traži model koji je najjednostavniji (najmanje postavke parametra podešavanja) i kod kojeg su performanse unutar jedne standardne greške od prethodno određene najbolje vrijednosti, konkretno tačnost ne manja od 97.50% (97.95% - 0.45%) za SMO model i ne manja od 98.53% (98.85%-0.32%) za RandomForest model. Primjenom ovog postupka, biće izabran SMO model sa vrijednošću 2 za c parametar i RandomForest model sa vrijednošću parametra numIterations koja je jednaka 900.

Moguća opcija za odabir je i izbor unutar određene tolerancije u odnosu na najbolje postignutu numeričku vrijednost (kao što se može vidjeti, sa profila u Tabeli 1, radi se o procjenjenoj tačnosti 97.95%). Tolerancija, procenat smanjenja performansi se može dobiti kao količnik, razlike vrijednosti performanse i numerički najbolje vrijednosti, i numerički najbolje vrijednosti (u skladu sa tim za najbolju vrijednost 97.95% tolerancija je jednaka 0). Nadalje, bira se zavisno od prihvatljivosti gubitka u tačnosti, u zamjenu za jednostavniji model.

Uzimajući u obzir vrijednosti profila postignutih rezultata, kao i vrijeme izvršavanja koje značajno raste sa primjenom tehnikom ponovnog uzorkovanja i porastom vrijednosti parametara (za neke od viših vrijednosti nisu se mogli završiti eksperimenti na korištenoj računarskoj opremi), uvažavajući opštu tendenciju prednosti jednostavnih modela sa prihvatljivim performansama u odnosu na kompleksnije i složenije modele (korištenjem SMO klasifikatora, za manje vrijednosti parametra c na primjer), finalno setovanje za potrebe izgradnje prediktivnog modela održavanja avionskog motora korištenjem vektora podrške za parametar c iznos 2, a vrijednost parametra broj stabala/numIterations iznosi 900 za model slučajna šuma.

Ako se za odabir između ta dva modela iskoristi površina ispod ROC krive, data za svaki od modela na Sl.4., koja sa vrijednostima od 0,9399 za SMO i 0.9981 za RandomForest ukazuje na izvanredno razdvajanje, prednost je jasno na strani modela izgrađenog korištenjem RandomForest algoritma.

U radu je prikazan, jedan od niza koraka, jedan od mnogih segmenta, na putu izgradnje prediktivnog modela održavanja avionskog motora, koji se ogleda u primjeni strategija koje u krajnjem treba da daju doprinos pouzdanosti. Prikazani postupak, ilustrovan za parametre SVM i RandomForest modela može se na adekvatan način primjeniti i na druge modele, za pripadajuće vrijednosti parametara podešavanja tih modela, znajući i uvažavajući prirodu paramet(a)ra pripadajućeg modela.

LITERATURA

- [1] P. Jahnke, "Machine Learning Approaches for Failure Type Detection and Predictive Maintenance," Master Thesis, 2015.
- [2] P. Domingos, "A Few Useful Things to Know about Machine Learning," Magazine Communications of the ACM, vol. 55, iss. 10, pp. 78-87, 2012.
- [3] J. Zarei, M.A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," Mechatronics, vol. 24, iss. 2, pp.151–157, 2014.
- [4] T.Y. Wu, J. Chen, and C.X. Wang, "Characterization of gear faults in variable rotating speed using hilbert-huang transform and instantaneous dimensionless frequency normalization," Mechanical Systems and Signal Processing, vol.30, pp.103–122, 2012.
- [5] H. Rødseth, and P. Schjølberg, "Data-driven Predictive Maintenance for Green Manufacturing," NTNU, Trondheim, Norway, International Workshop of Advanced Manufacturing and Automation (IWAMA 2016), [Online]. Dostupno: <http://www.atlantis-press.com>
- W. Liao and Y. Wang, "Data-driven Machinery Prognostics Approach using in a Predictive Maintenance Model," 2013., [Online]. Dostupno: <http://www.jcomputers.us>
- [6] M. Lichman, UCI Machine Learning Repository Irvine, University of California, 2013. [Online]. Dostupno: <http://archive.ics.uci.edu/ml>
- [7] NASA Ames Research Center, NASA Ames Prognostics Data Repository, 2008 ([Online]. Dostupno: <http://ti.arc.nasa.gov/project/prognostic-data-repository>
- [8] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-Failure," Ist PHM08, 2008. [Online]. Dostupno: <https://www.researchgate.net>
- [9] O. F. Eker, F. Camci, and I. K. Jennions, "Major Challenges in Prognostics: Study on Benchmarking Prognostics Datasets," PHME12, 2012., [Online]. Dostupno: https://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2012/phmce_12_004.pdf
- [10] E. Yakovleva, and P. Erofeev, "Data-driven Models for Run-to-failure Time Prediction for Aircraft Engines," Information Technologies and Systems (ITaS) 2015, [Online]. Dostupno: <http://itas2015.iitp.ru/>

- [11] M. Cottrell, P. Gaubert, C. Eloy, D. François, G. Hallaux, J. Lacaille, and M. Verleysen, "Fault prediction in aircraft engines using SelfOrganizing Maps," *Advances in Self-Organizing Maps*, pp. 37-44, 2009., [Online]. Dostupno: <https://arxiv.org/ftp/arxiv/papers/0907/0907.1368.pdf>
- [12] C. Brasco, N. Eklund, M. Shah, and D. Marthaler, "Predictive Modeling of High-Bypass Turbofan Engine Deterioration," *Annual Conference of the PHM Society* 2013. [Online]. Dostupno: http://www.mohakshah.com/Site/Publications_files/phm2013-2.pdf
- [13] D. Feng, M. Xiao, Y. Liu, H. Song, Z. Yang, and L. Zhang, "A kernel principal component analysis-based degradation model and remaining useful life estimation for the turbofan engine," *Advances in Mechanical Engineering*, vol. 8 pp. 1-13, 2016. [Online]. Dostupno: <http://journals.sagepub.com/doi/full/10.1177/1687814016650169>
- [14] A.K. Jain, P. Kundu, B.K. Lad, "Prediction of Remaining Useful Life of an Aircraft Engine under Unknown Initial Wear," *AIMTDR* 2014, 2014, [Online]. Dostupno: <http://www.iitg.ernet.in/aimtdr2014/PROCEEDINGS/papers/494.pdf>
- [15] Y.S. Chati, and H. Balakrishnan, "A Gaussian Process Regression Approach to Model Aircraft Engine Fuel Flow Rate," *ICCPS* 2017, 2017. [Online]. Dostupno: <http://www.mit.edu/~hamsa/pubs/ChatiBalakrishnanICCPS2017.pdf>
- [16] J.L. Loyer, E. Henriques, and S. Wiseall, "Comparison of binary classifiers for data-driven prognosis of jet engines health," *European Conference of the Prognostics and Health Management Society* 2014
- [17] Air France Industries KLM Engineering, & Maintenance Prognos: the Predictive Aircraft Maintenance tool. Air France Industries KLM Engineering & Maintenance, 16 June 2017
- [18] Pratt & Whitney's 'Big Data' Projects Advancing Analytics Efforts in Aftermarket, JUNE 5, 2015. [Online]. Dostupno: <http://www.pw.utc.com/News/Story/20150605-1400/2015/AI%20Categories>
- [19] H. Canaday, Sharing Data For Predictive Maintenance, Mar 9, 2017. [Online]. Dostupno: <http://aviationweek.com/aviation-maintenance-and-support-software/sharing-data-predictive-maintenance>
- [20] A.S. Sanghavi, "Man vs. Machine: Aviation's Predictive Maintenance Challenge," 2016, [Online]. Dostupno: <http://www.aviationpros.com/article/12236185/man-versus-machine-aviations-predictive-maintenance-challenge>
- [21] K. Max, and J. Kjell, *Applied Predictive Modeling*, Published May 17th 2013 by Springer
- [22] O. Janković, "Inženjering karakteristika u kontekstu predikcije korištenjem binarne klasifikacije," *XXIII naučna i biznis konferencija YU INFO* 2017, Zbornik radova, str.46-51, 2017
- [23] O. Janković, "Implementacija modela prediktivnog održavanja korištenjem binarne klasifikacije," *INFOTEH* 2017, Vol. 16, str.576-581, 2017

ABSTRACT

Regardless of the applied machine learning techniques, when building prediction models based on data, the imperative of each modeler is reliable prediction. The predictability of prediction is basically based on the ability of a predictive model to achieve prediction performance on the data on which the model is trained, and to confirm that on previously unknown test data, with the aim of ultimately reproducing it on the data used after the model application. The paper presents the approach based on the confidence of the predictive model reliability contribution, implemented for the reliability of prediction of the predictive model for aircraft maintenance, by contributing to the reliability of the used binary classification classifiers.

The contribution of reliability prediction for the predictive model for aircraft engine maintenance

Olivera Janković, Đorđe Babić