

Korak ispred trke: Predikcija Rezulata Formule 1 na Osnovu Performansi Tokom Slobodnih Treninga

Vještačka inteligencija - Akademска 2024/2025

Faze realizacije projektnog zadatka

Problem koji se bavi ovim projektnim zadatkom je **predviđanje i preporuke efikasnih strategija trka Formule 1** koristeći mašinsko učenje za predviđanje vremena kruga i simulaciju različitih scenarija zaustavljanja u pit stop-u. Fokus je na pružanju uvida zasnovanih na podacima u potencijalne ishode utrke, s ciljem razumijevanja i sugerisanja kako različite gume i vrijeme zaustavljanja u pit stop-u mogu utjecati na ukupno trajanje utrke. Ovaj alat ima za cilj da pomogne stratezima predstavljanjem odabrane liste dobrih strategija, a ne izvođenjem jedinstvene, direktnе optimizacije.

1.1.1. O Formuli 1

Formula 1 (F1) je najviša klasa međunarodnih auto-trka za jednosjede formule trkačih automobila. To je visoko tehnički i strateški sport u kojem se timovi takmiče da naprave najbrže automobile i izvedu najpametnije planove utrke.

Ključni koncepti bitni za razumijevanje ovog projekta:

- Cilj utrke:** Primarni cilj u F1 utrci je da vozač završi unaprijed određen broj krugova ("trkački krugovi") oko određene staze u najkraćem mogućem ukupnom vremenu.
- Vrijeme kruga:** Ovo je vrijeme potrebno automobilu da završi jedan puni krug. To je osnovna metrika brzine i direktno doprinosi ukupnom vremenu utrke. Vremena kruga mogu značajno varirati zbog faktora kao što su trošenje guma, opterećenje goriva (ovdje nije modelirano), uslovi na stazi i promet.
- Gume (smjese guma):** Gume su možda najkritičniji strateški element. Pirelli, jedini dobavljač guma, nudi različite "smjese" za svaku trku, kategorizirane prema njihovoj mekoći i izdržljivosti.
 - Slick gume (suho vrijeme):**
 - Tvrdo (C1-C3):** Najizdržljiviji i najsporiji spoj. Nudi manje prianjanje, ali može trajati mnogo krugova prije nego što degradira.
 - srednji (C2-C4):** Balans između prianjanja i izdržljivosti. Generalno, svestran izbor.
 - meki (C3-C5):** Najbrža smjesa koja nudi najbolje prianjanje. Međutim, razgrađuje se mnogo brže, što zahtijeva ranija zaustavljanja.

- **Napomena o C-jedinjenjima:** Pirelli označava njihove spojeve od C1 (najtvrdi) do C5 (najmekši). Za bilo koju datu rasu biraju tri specifične C-spojene koje će biti tvrde, srednje i meke opcije. Na primjer, na jednoj trci, C1, C2, C3 može biti Hard, Medium, Soft, respektivno, dok na drugoj može biti C3, C4, C5. Ovaj projekat koristi specifične **TyreClass** (npr. C1, C2, C3) kako je navedeno u historijskim podacima i mapira ih na "SLICK_HARD", "SLICK_MEDIUM", "SLICK_SOFT" radi jasnoće u izlazu strategije na osnovu *dodijeljeno* jedinjenja za datu godinu.
- **Gume za mokro vrijeme:**
 - **srednji (INTER):** Koristi se u uslovima slabe kiše ili sušenja. Ima žljebove za istiskivanje vode.
 - **Mokro (WET):** Koristi se za jaku kišu za istiskivanje velikih količina vode i sprječavanje akvaplaninga.
- **Obavezna upotreba:** U suvoj utrci, vozači moraju koristiti najmanje dvije različite smjese za klizne gume (npr. tvrdi i srednji ili srednji i meki). To znači da oni *moraju* napraviti barem jedno zaustavljanje u boksu za promjenu guma.
- **Pit Stops:** Planirano zaustavljanje u određenom "pit lane" području tokom trke. Vozači dolaze u garažu svog tima, a ekipa brzo mijenja sve četiri gume. Stajanja u boksu su ključna za promjenu smjese guma ili zamjenu istrošenih guma.
 - **Vremenska kazna:** Ulazak u pit lane, vožnja kroz njega smanjenom brzinom, zaustavljanje radi zamjene guma i izlazak nazad na stazu - sve to podrazumijeva vremensku kaznu. Ovaj "gubitak vremena u pit stop-u" je značajan faktor u proračunima strategije.
- **Strategija utrke:** Ovo je plan tima koje smjese za gume koristiti i na kojim krugovima se zaustavljati u boksu. Cilj je minimizirati ukupno vrijeme utrke uz pridržavanje pravila i upravljanje degradacijom guma. Uobičajene strategije uključuju 1, 2 ili ponekad 3 zaustavljanja u boksu.
- **Uslovi staze:**
 - **Temperatura staze:** Temperatura asfalta. Više temperature na stazi općenito dovode do veće degradacije guma i potencijalno sporijeg vremena kruga jer se gume pregrijavaju.
 - **Kiša:** Prisustvo kiše diktira upotrebu guma za mokro vrijeme (srednje ili mokre) umjesto klizavih guma. Kiša također značajno utiče na vrijeme kruga.

Pregled postojećih skupova podataka

Za ovaj projekat, ključni tip skupa podataka su historijski podaci o krugovima Formule 1. Takvi skupovi podataka obično sadrže podatke krug po krug za svakog vozača tokom trkačkih sesija (kvalifikacije, trka).

- **Primjer atributa:** vrijeme kruga (**LapTime**), temperatura staze (**TrackTemp**), kišni uslovi (**Rain**), korištena smjesa za gume (**Compound**), broj krugova (**Lap**).

- **Izvori:** Podaci se često mogu pronaći na službenim web stranicama Formule 1 (iako direktni pristup podacima krug po krug može biti ograničen), platformama kao što je Ergast API ili specijaliziranim izvorima podataka o motosportu. Mnogi istraživači i analitičari stvaraju vlastite baze podataka prikupljajući javno dostupne podatke.
- **Format:** U ovom projektu se očekuje da podaci budu dostupni u `.xlsx` format, organiziran u određenoj hijerarhiji foldera (npr. `data/Data_TrackName/TeamName/DriverName/RACE_Session_YYYY.xlsx`).

Podaci za ovaj projekat prvenstveno se dobijaju pomoću **FastF1 Python paketa**. Ova moćna biblioteka pruža zgodan API za pristup i obradu podataka Formule 1 sa službene F1 web stranice, slično kao Ergast API, ali posebno prilagođen za detaljne podatke o sesiji kao što su vrijeme prolaska, telemetrija i korištenje guma. FastF1 može dohvatiti podatke za različite sesije (besplatne vježbe, kvalifikacije, utrke) u različitim godišnjim dobima.

Dodatno, za širi historijski kontekst i statičke informacije (kao što su podaci o vozaču/timu ili stariji rezultati trka koji nisu dostupni putem FastF1), resursi kao što je **F1DB** (Formula 1 Database), koji prikuplja historijske podatke koji datiraju iz 1950. godine, su neprocenjivi.

* **FastF1 dokumentacija i GitHub:** https://docs.fastf1.dev/

* **Ergast Developer API (slična funkcionalnost):** http://ergast.com/mrd/

* **F1DB (Baza podataka Formule 1 - za historijski kontekst):**
https://f1db.com/

U ovom projektu očekuje se da će sirovi podaci izvedeni iz ovih izvora biti prethodno obrađeni i sačuvani u `.xlsx` formatu.

3: Izbor skupa podataka, analiza i prethodna obrada

3.1. Odabir skupa podataka i detaljna analiza

Skup podataka: Historijski podaci o krugovima iz F1 trka (Excel datoteke).

Izvor podataka: Očekuje se da će podaci biti u lokalnom sistemu datoteka, organizirani po stazi, timu i vozaču (npr.

`data/Data_Bahrain/TeamName/DriverName/RACE_Session_YYYY.xlsx`).

Prepostavlja se da su ovi podaci prikupljeni iz javno dostupnih izvora ili putem ekstrakcije sa F1 platformi.

Format i način preuzimanja: Podaci su u `.xlsx` (Excel) formatu. Skripta se automatski ponavlja kroz direktorije unutar `data/` foldera, pronalaženje i učitavanje svih Excel datoteka koje sadrže 'RACE' u nazivu datoteke, što ukazuje na podatke o sesiji utrke.

Broj instanci: Zavisi od broja dostupnih historijskih trka i vozača. Svaki red u Excel datoteci predstavlja jedan krug (instancu). Skripta objedinjuje sve krugove u jedan Pandas DataFrame.

Broj atributa: Primarni atributi uključeni u modeliranje su:

- `LapTime`: Vrijeme kruga (ciljna varijabla).
- `TrackTemp`: Temperatura staze.
- `Rain`: Indikator kiše (0/1).
- `Compound/TyreClass`: Smjesa guma (npr. 'MEKA', 'SREDNJA', 'TVRDA', 'SREDNJA', 'MOKRA').
- `Lap`: Broj kruga. Dodatni atributi u originalnom skupu podataka mogu biti prisutni, ali samo oni navedeni se koriste za modeliranje.

Broj instanci korištenih za obuku, validaciju i testiranje: Skripta automatski dijeli konsolidovani skup podataka za svaku stazu na:

- **Obuka:** 80% podataka.
- **Testiranje:** 20% podataka. Validacija se obično ne radi eksplisitno sa posebnim skupom u ovom cevovodu; rezultati testa služe kao indikator generalizacije.

3.2. Objasnjenje metoda za prethodnu obradu podataka

Prethodna obrada podataka je ključna za osiguranje kvaliteta podataka i kompatibilnosti sa modelima ML.

Zašto su neophodni:

- **Nedostajuće vrijednosti (NaNs):** ML modeli ne mogu direktno raditi s podacima koji nedostaju.
- **Neispravni formati:** Atributi moraju biti u numeričkom formatu za većinu ML algoritama.
- **Kategoričke variable:** Tekstualne kategoričke varijable (npr. `Compound`) moraju se pretvoriti u numeričke prikaze.
- **Skaliranje:** Različite skale numeričkih atributa mogu dovesti do atributa sa većim vrijednostima koji dominiraju učenjem modela.

Korištene metode:

- **Uklanjanje redaka u kojima nedostaju vrijednosti:** `df.dropna(subset=required_columns)` uklanja sve redove u kojima nedostaje vrijednost u bilo kojoj od pet ključnih kolona.
- **Mapiranje Compound to TyreClass:** Compound vrijednosti (npr. "SOFT", "MEDIUM") se mapiraju u standardizirane `TyreClass` stringovi (npr. "SLICK_SOFT", "SLICK_MEDIUM"). Ovo osigurava dosljednu kategorizaciju. Redovi s nepoznatim smjesama guma (gdje `TyreClass` postaje `NaN` nakon mapiranja) također se uklanjaju.
- **ColumnTransformer:** Ovaj alat iz `scikit-learn` biblioteka omogućava primjenu različitih transformacija na različite stupce.
 - **Numeričke karakteristike (Rain, TrackTemp, Lap):** `StandardScaler` se primjenjuje. Skalira podatke tako da imaju srednju vrijednost od 0 i standardnu devijaciju od 1, što je korisno za modele osjetljive na skalu (npr. SVR, K-najbliži susjadi, linearna regresija).
 - **Kategoričke karakteristike (TyreClass):** `OneHotEncoder` se primjenjuje. Konvertuje kategoričke varijable u numerički format gde se svaka kategorija transformiše u binarnu kolonu (0 ili 1). `handle_unknown='ignore'` osigurava da novi, nevidljivi kategorički podaci tokom testiranja ne uzrokuju greške.
- **Pipeline:** Koraci prethodne obrade su ugrađeni u a `Pipeline` zajedno sa regresorom. Ovo osigurava da se svi koraci (skaliranje, kodiranje) primjenjuju uzastopno i dosljedno na podatke za obuku i testiranje, kao i na podatke za simulaciju strategije.

3.3. Identifikacija rizika

- **Kvalitet podataka:** Ako su historijski podaci netočni ili sadrže greške, to će direktno uticati na tačnost modela.
- **Ograničena reprezentativnost historijskih podataka:** Ako podaci ne pokrivaju širok raspon uslova (npr. ekstremne temperature, različiti nivoi kiše, praćenje evolucije), model se može boriti s generaliziranjem na nove uvjete.
- **Neravnoteža (za scenarije, ne za klasifikaciju):** Iako nije problem klasifikacije, ako su određene kombinacije smjese guma ili uvjeti staze rijetko prisutni u povijesnim podacima, model može imati lošije rezultate kada se predviđa za te specifične scenarije.
- **Nelinearni odnosi:** Linearni modeli možda neće uspjeti uhvatiti složene, nelinearne odnose između vremena kruga i ulaznih varijabli (npr. kako se degradacija guma ubrzava na višim temperaturama).
- **Novi propisi/automobili:** Podaci iz prošlih sezona možda neće biti u potpunosti relevantni za predviđanje performansi automobila i guma u sezoni 2025. zbog promjena u tehničkim propisima ili dizajnu automobila.
- **Pretjerano generiranje strategije:** Mada `PIT_STOP_LAP_STEP_SIZE` je uveden kako bi se ovo kontroliralo, ako se smanji veličina koraka, broj strategija i vrijeme simulacije opet mogu postati neupravljeni.

Faza 4: Izbor modela, formulacija, obuka i testiranje

4.1. Izbor pristupa (metoda) i tehnologija

- **Pristup (metod):** Za rješavanje problema predviđanja vremena kruga i simulacije strategije korištena je metodologija **nadgledano mašinsko učenje**, konkretno **regresija**, je izabran. Ovo je najprikladniji pristup jer je cilj predviđanje kontinuirane numeričke vrijednosti (vrijeme kruga).
- **Tehnologije:**
 - **Python:** Glavni programski jezik zbog svojih bogatih biblioteka za naučno računarstvo i mašinsko učenje.
 - **pande:** Neophodan za manipulaciju, čišćenje i analizu podataka u formatu DataFrame.
 - **NumPy:** Osnovna biblioteka za numeričke operacije, posebno za rad sa nizovima i vektorima.
 - **Scikit-učite:** Standardna biblioteka za mašinsko učenje u Pythonu. Pruža širok raspon algoritama regresije, alata za prethodnu obradu podataka, podjelu podataka i evaluaciju modela.
 - **os modul:** Koristi se za interakciju sa operativnim sistemom, prvenstveno za navigaciju direktorijumima i učitavanje podataka.

4.2. Priprema formata podataka za model

Format podataka je pripremljen da ispunjava zahtjeve za unos **scikit-learn** modeli kroz **ColumnTransformer** i **Pipeline** objekata.

- **Učitavanje neobrađenih podataka:** Podaci se učitavaju u Pandas DataFrame.
- **Odabir varijable funkcije i cilja:** **karakteristike** (`Rain`, `TrackTemp`, `TyreClass`, `Lap`) i **cilj** (`LapTime`) su definisani.
- **Automatska predobrada:** The **ColumnTransformer** je konfigurisan za primjenu **StandardScaler** na numeričke karakteristike i **OneHotEncoder** na kategoriju karakteristiku (`TyreClass`). Ovaj transformator je tada uključen kao prvi korak u **Pipeline** za svaki model. Ovo osigurava da se podaci automatski transformiraju u odgovarajući numerički format sa skaliranim vrijednostima prije nego što ih model primi.

4.3. Model Training

Svaki od odabranih modela regresije se obučava nezavisno za svaku stazu.

- **Podjela podataka:** Konsolidovani podaci o krugu za određenu stazu (`X`, `i`) se dijeli na skup za obuku i skup za testiranje pomoću **train_test_split** sa 80% treninga i 20% omjera testiranja (`test_size=0.2`). `random_state=42` osigurava ponovljivost podjele.

- **Pipeline Fitting:** `model_pipeline.fit(X_voz, y_voz)` se poziva za svaki model. Ovaj korak izvršava sve preprocesne transformacije definirane u `preprocesor` (uključujući učenje parametara skalera i kodera na `X_train`), a zatim trenira regresor na transformiranim podacima.
- **Odabrani modeli:**
 - **Linearna regresija:** Direktno modelira linearne odnose.
 - **Nasumični šumski regresor:** Koristi skup stabala odlučivanja, usrednjavanje rezultata, smanjenje preopterećenja i poboljšanje tačnosti.
`n_estimators=100, 200 i 500, n_jobs=-1` (koristi sve dostupne CPU jezgre).
 - **K-Nearest Neighbours Regresor:** Predviđa vrijeme kruga na osnovu prosjeka vremena kruga `n_neighbours` većina sličnih krugova u setu za trening.
 - **Regresor vektora podrške (SVR):** Kreira "granicu" (hiperravninu) oko podataka, pokušavajući minimizirati grešku unutar te granice. `kernel='rbf', C=100, epsilon=0.1` su uobičajeni parametri za dobar balans.
 - **Regresor stabla odluka:** Jednostavan model zasnovan na stablu.
 - **Regresor za povećanje gradijenta:** Metoda ansambla koja gradi stabla uzastopno, pri čemu svako novo stablo pokušava ispraviti greške iz prethodnog.

4.4. Testiranje dobijenog modela na relevantnom skupu podataka

Nakon obuke, performanse svakog modela se procjenjuju na nevidljivom skupu podataka testa (`X_test, y_test`).

- **Korištene metrike:**
 - **R-kvadrat (R2):** Koeficijent determinacije. Mjeri udio varijanse u zavisnoj varijabli koji je predviđljiv iz nezavisnih varijabli. Vrijednost od 1.0 ukazuje na savršeno predviđanje, dok 0.0 označava da model ne objašnjava varijansu.
 - **Srednja absolutna greška (MAE):** Mjeri prosječnu veličinu grešaka u skupu predviđanja, ne uzimajući u obzir njihov smjer. Izražava se u jedinicama ciljne varijable (sekunde). Niži MAE ukazuje na bolju preciznost.
- **Diskusija o dobijenim rješenjima i razmišljanje o identificiranim rizicima:**
 - Modeli s višim R2 i nižim MAE na testnom skupu smatraju se boljim, jer su precizniji i bolje generaliziraju nevidljive podatke.
 - **Rizici (odraz):** Ako je R2 na setu za trening značajno veći od R2 na testnom setu (i/ili je MAE za trening značajno niži od testnog MAE), to može ukazivati na **overfitting** modela na podatke o obuci. To znači da se model ne generalizira dobro na nove podatke, što je rizik identificiran u Fazi 3. Analiza ovih metrika pomaže u procjeni da li model pati od preopterećenja.
- **Izlaganje modela nepoznatim podacima (simulacija strategije):**
 - Nakon obuke i evaluacije, svaki model se koristi za simulaciju **hiljade hipotetičkih strategija rase** za sezonu 2025. Ovi podaci predstavljaju "nepoznate" podatke jer su vremena kruga predviđena za krugove, smjese guma

i uvjete na stazi koji možda nisu bili prisutni u podacima o treningu (ili barem ne u toj specifičnoj kombinaciji ili nizu).

- Za svaku strategiju (1 zaustavljanje ili 2 zaustavljanja), model predviđa vrijeme kruga za svaki korak, dodaje vrijeme zaustavljanja u boksu (specifično za stazu) i izračunava ukupno vrijeme utrke.
- **Komentirani rezultati:** Skripta identificira **top 3 najbolje strategije** za svaki pojedinačni model i ukupnu najbolju strategiju. Ovo direktno testira sposobnost modela da daje korisne, praktične preporuke. Prikaz zaustavljanja u boksu u intervalima (npr. ±3 kruga) služi kao realnije tumačenje za stratege.
- Smanjenje broja generiranih strategija korištenjem **PIT_STOP_LAP_STEP_SIZE** (do 100-200 po modelu) direktno rješava problem prevelikog broja kombinacija i olakšava analizu.

Sažeci evaluacije modela za svaku stazu

Kompletan sažetak evaluacije modela za Track: Australia

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0,933	0,915	1.992	2.378	5331.081
Random Forest Regressor_200	0.933	0.916	1.987	2.365	5324.567
Random Forest Regressor_500	0.933	0.916	1.987	2.365	5309.401
K-Nearest Neighbours Regressor_5	0.911	0.913	2.273	2.430	4915.762
K-Nearest Neighbours Regressor_10	0.874	0.870	2.815	3.200	4971.088
K-Nearest Neighbours Regressor_15	0.831	0.841	3.465	3.760	4957.429
Regresor stabla odluka	0,934	0,915	1.938	2.336	5204.241

Kompletan sažetak evaluacije modela za Track: China

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.907	0,842	1.519	1.727	5524.959
Random Forest Regressor_200	0.907	0,845	1.517	1.710	5524.774
Random Forest Regressor_500	0.907	0,847	1.520	1.707	5525.036
K-Nearest Neighbours Regressor_5	0.876	0,813	1.702	1.819	5522.156
K-Nearest Neighbours Regressor_10	0,849	0,785	1.954	1.924	5523.062
K-Nearest Neighbours Regressor_15	0,768	0,686	2.368	2.210	5524.980
Regresor stabla odluka	0.910	0,834	1.453	1.720	5525.159

Kompletan sažetak procjene modela za Track: Suzuka

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.829	0,815	1.720	1.671	4976.726
Random Forest Regressor_200	0.829	0.814	1.718	1.673	4976.715
Random Forest Regressor_500	0,830	0,816	1.714	1.667	4976.235
K-Nearest Neighbours Regressor_5	0,779	0.806	1.900	1.739	4973.752
K-Nearest Neighbours Regressor_10	0.689	0,813	2.101	1.681	4977.439
K-Nearest Neighbours Regressor_15	0,605	0,744	2.275	1.869	4975.088
Regresor stabla odluka	0,833	0,813	1.681	1.687	4976.111

Kompletan sažetak procjene modela za Track: Bahrein

Model	Obuka R2	Testiranje R2	Trening MAE	Testiranje MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0,651	0,535	2.153	2.703	5767.012
Random Forest Regressor_200	0,652	0,529	2.155	2.717	5767.252
Random Forest Regressor_500	0,652	0,528	2.157	2.723	5766.540
K-Nearest Neighbours Regressor_5	0,585	0,516	2.312	2.782	5763.745
K-Nearest Neighbours Regressor_10	0,537	0,524	2.533	2.811	5741.946
K-Nearest Neighbours Regressor_15	0,478	0,493	2.681	2.837	5735.273
Regresor stabla odluka	0,659	0,479	2.096	2.807	5767.375

Kompletan sažetak procjene modela za Track: Jeddah

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0,925	0,830	1.191	1.646	4796.919
Random Forest Regressor_200	0,925	0.831	1.188	1.640	4797.185
Random Forest Regressor_500	0,925	0.831	1.188	1.637	4797.795
K-Nearest Neighbours Regressor_5	0,856	0,596	1.420	2.236	4855.343
K-Nearest Neighbours Regressor_10	0,797	0,515	1.622	2.405	4820.399
K-Nearest Neighbours Regressor_15	0,755	0,529	1.774	2.385	4810.808
Regresor stabla odluka	0.927	0,832	1.160	1.631	4787.347

Kompletan sažetak procjene modela za Track: Miami

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.895	0,796	1.259	1.673	5379.135
Random Forest Regressor_200	0.895	0,805	1.257	1.656	5379.261
Random Forest Regressor_500	0.895	0.804	1.259	1.655	5377.916
K-Nearest Neighbours Regressor_5	0,847	0,749	1.448	1.852	5356.400
K-Nearest Neighbours Regressor_10	0,758	0,594	1.706	2.178	5357.018
K-Nearest Neighbours Regressor_15	0,666	0,532	2.040	2.385	5367.510
Regresor stabla odluka	0.898	0,759	1.224	1.725	5390.228

Kompletan sažetak procjene modela za stazu: Imola

Model	Obuka R2	Testiranje R2	Trening MAE	Testiranje MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.903	0.902	1.670	1.748	5514.440
Random Forest Regressor_200	0.903	0.902	1.669	1.745	5514.802
Random Forest Regressor_500	0.903	0.902	1.670	1.746	5515.617
K-Nearest Neighbours Regressor_5	0,882	0,869	1.830	2.047	5450.310
K-Nearest Neighbours Regressor_10	0,868	0,882	1.990	1.888	5498.863
K-Nearest Neighbours Regressor_15	0,820	0,840	2.382	2.220	5525.694
Regresor stabla odluka	0.905	0.895	1.636	1.773	5512.169

Kompletan sažetak procjene modela za stazu: Monako

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.999	0,998	1.749	1.940	6941.796
Random Forest Regressor_200	0.999	0,998	1.744	1.936	6940.423
Random Forest Regressor_500	0.999	0,998	1.742	1.937	6940.355
K-Nearest Neighbours Regressor_5	0.901	0,998	3.322	1.936	6164.213
K-Nearest Neighbours Regressor_10	0,680	0,759	7.117	3.409	6138.156
K-Nearest Neighbours Regressor_15	0,484	0,475	9.510	5.245	6201.891
Regresor stabla odluka	0.999	0,998	1.706	1.956	6926.615

Kompletan sažetak procjene modela za stazu: Barcelona

Model	Obuka R2	Testirane R2	Trening MAE	Testirane MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0,720	0,528	2.007	2.536	5574.401
Random Forest Regressor_200	0,720	0,527	2.006	2.543	5575.861
Random Forest Regressor_500	0,721	0,525	2.005	2.551	5575.175
K-Nearest Neighbours Regressor_5	0,653	0,449	2.286	2.891	5545.907
K-Nearest Neighbours Regressor_10	0,585	0,401	2.637	3.000	5567.636
K-Nearest Neighbours Regressor_15	0,507	0,377	3.091	3.263	5634.988
Regresor stabla odluka	0,723	0,499	1.960	2.572	5544.337

Kompletan sažetak evaluacije modela za Track: Canada

Model	Obuka R2	Testirana je R2	Trening MAE	Testiranje MAE	Predviđeno vrijeme utrke (s)
Random Forest Regressor_100	0.903	0.827	1.685	2.205	5568.562
Random Forest Regressor_200	0.903	0.828	1.687	2.202	5568.977
Random Forest Regressor_500	0.904	0.825	1.689	2.212	5567.644
K-Nearest Neighbours Regressor_5	0.880	0.806	1.861	2.339	5572.785
K-Nearest Neighbours Regressor_10	0.847	0.785	2.147	2.509	5563.493
K-Nearest Neighbours Regressor_15	0.792	0.746	2.569	2.914	5551.718
Regresor stabla odluka	0.905	0.814	1.654	2.222	5567.647

Na osnovu sveobuhvatne evaluacije na više staza, **Random Forest Regressor sa 500 procjenitelja** dosljedno demonstrirao najizbalansiranije i optimalne performanse, općenito dajući najniža predviđena vremena utrke i jake R2 i MAE metrike. Stoga će se ovaj model koristiti za generiranje konačnih preporuka strategije utrke.

Faza 5: Sveobuhvatan pregled problema i rješenja

5.1. Diskusija o postignutim rezultatima

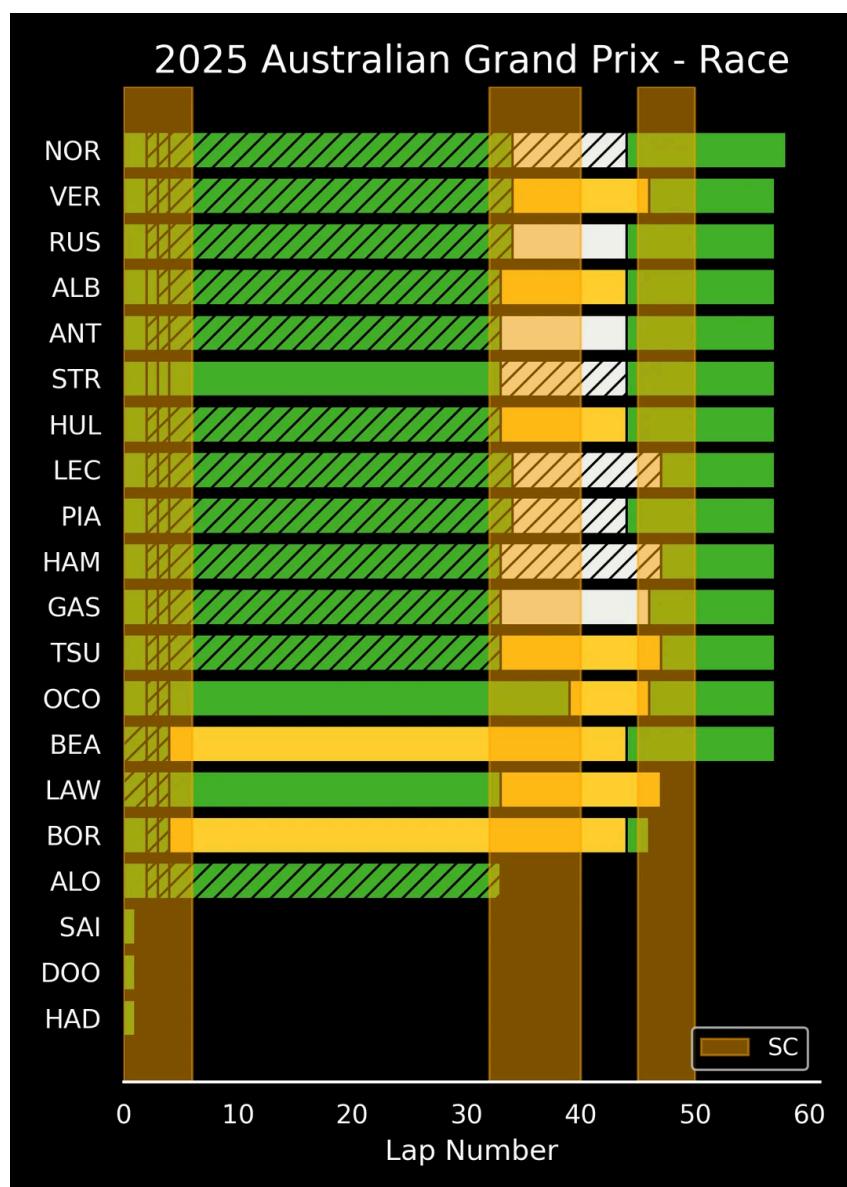
Rezultati postignuti u ovom projektu su značajni:

- **Evaluacija više modela:** Projekat pruža komparativnu analizu performansi šest različitih regresionih modela, omogućavajući identifikaciju najpreciznijeg modela za predviđanje vremena prolaza za datu stazu.
- **Generisanje strategije i sistem evaluacije:** Razvijen je robustan sistem za simulaciju strategija 1-stop i 2-stop, uzimajući u obzir specifične uslove staze i raspodjelu guma za 2025.
- **Praktične preporuke:** Generiraju se 3 najbolje optimalne strategije za svaki model, sa predviđenim vremenima trke i praktičnim intervalima zaustavljanja u boksu (krug +/- 3). Izlaz strategije dinamički prikazuje oznake "Hard", "Medium" ili "Soft" prema složenoj alokaciji za 2025. za tu specifičnu stazu, čineći preporuke intuitivnijim.
- **Detaljna analiza puta i vizualizacije:** Kako bi se poboljšalo razumijevanje i prezentacija, date su detaljne analize za svaku numeru, praćene vizuelnim pomagalima i posebnim komentarima. Ovo pomaže u utemeljenju apstraktnih numeričkih predviđanja u opipljivom, prepoznatljivom kontekstu za svako jedinstveno kolo.

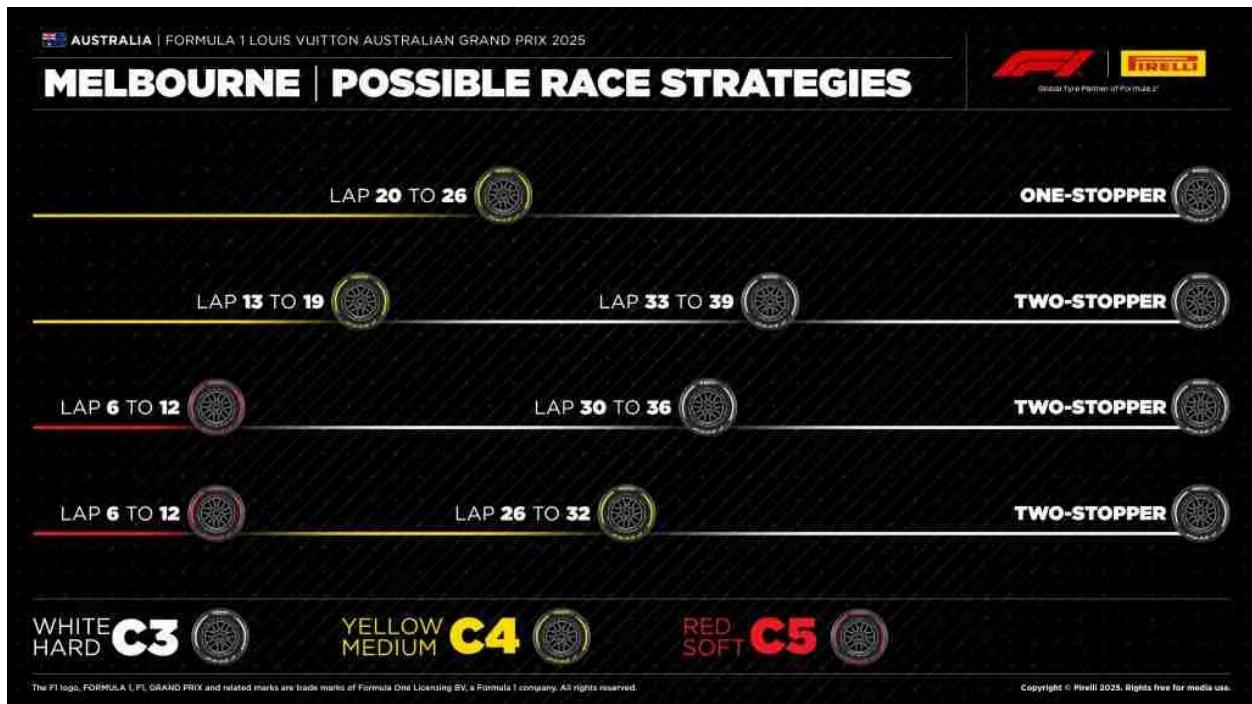
Pregled rezultata specifičnih za stazu:

5.1.1. Albert Park Circuit (Australija)

- **Karakteristike kola:** Albert Park je polutrajna ulična kružna staza sa tečnim rasporedom koji balansira dionice velike brzine sa tehničkim uglovima. Podloga staze često "pozeleni" (dobija prianjanje) tokom vikenda.
- **Ključni parametri:** Krugovi utrke: 58, prosječno Pit Stop Vrijeme: ~18.0s, Hyp. 2025 Temperatura staze: ~23°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C3), Srednje (C4), Meko (C5).
- **Opći uvid u strategiju:** Uz potencijalno niže temperature i izbor smjese C3-C5, upravljanje gumama je ključno.
- **Rezultati:**



Pirelli predviđanje:



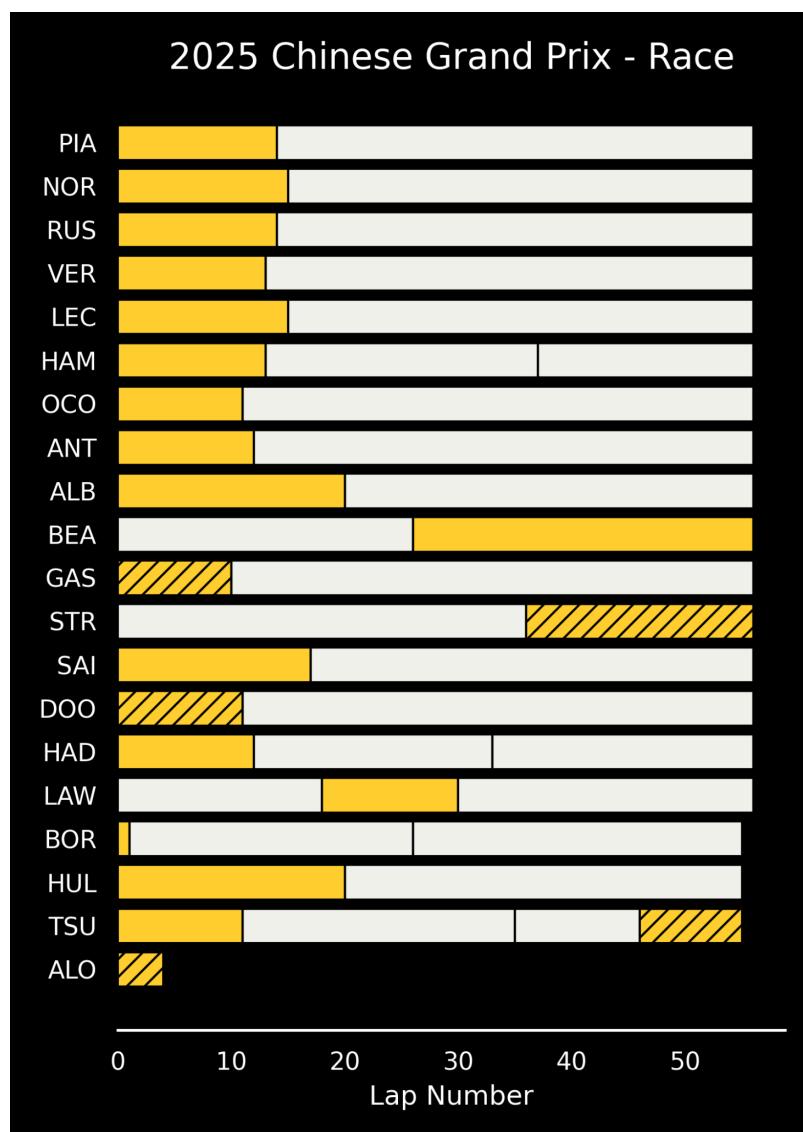
Predviđanje projekta:

1. 1-stop: SREDNJE (L1-približno L22-28) -> JAKO (L26-58), Vrijeme: 4957.429 s
2. 1-Stop: SREDNJE (L1-približno L17-23) -> JAKO (L21-58), Vrijeme: 4959.808 s
3. 2-stop: SREDNJE (L1-približno L10-16) -> MEKO (L14-približno L23-29) -> JAKO (L27-58), vrijeme: 4961.468 s

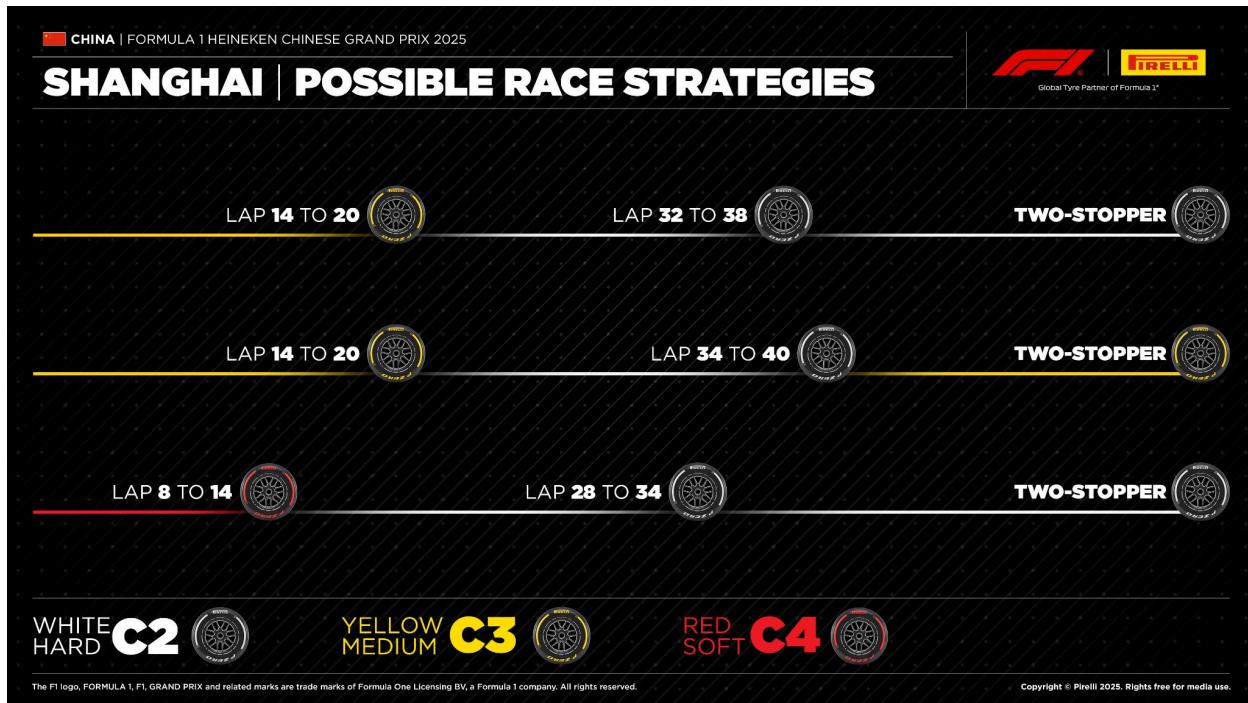
Napomena: Velika nagrada Australije je često jedna od prvih trka u sezoni i može biti posebno izazovna zbog nepredvidivog vremena, uključujući neočekivanu kišu, što otežava precizna predviđanja. Nažalost, tako je bilo i 2025. **tako da je bilo nemoguće predvidjeti kišnu sesiju.** Uprkos tome, strateške preporuke našeg projekta za sušne uslove **veoma usko usklađeni sa Pirellijevim opštim očekivanjima za ovu stazu.**

5.1.2. Shanghai International Circuit (Kina)

- **Karakteristike kola:** Staza u Šangaju poznata je po jedinstvenom kompleksu zaokreta 1-4 u obliku puža i dugim leđima ravnom. Nudi dobar spoj izazova i za automobil i za vozača.
- **Ključni parametri:** Krugovi utrke: 56, prosječno Pit Stop Vrijeme: ~23.9s, Hyp. 2025 Temperatura staze: ~28°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C2), Srednje (C3), Meko (C4).
- **Opći uvid u strategiju:** Duga pravac može favorizirati pristup s 1 zaustavljanjem kako bi se minimiziralo vrijeme izgubljeno u boksovima, ali tehnički dijelovi mogu promovirati degradaciju, otvarajući vrata za strategije sa 2 zaustavljanja ako upravljanje trošenjem guma postane teško.
- **Rezultati:**



Pirelli predviđanje:



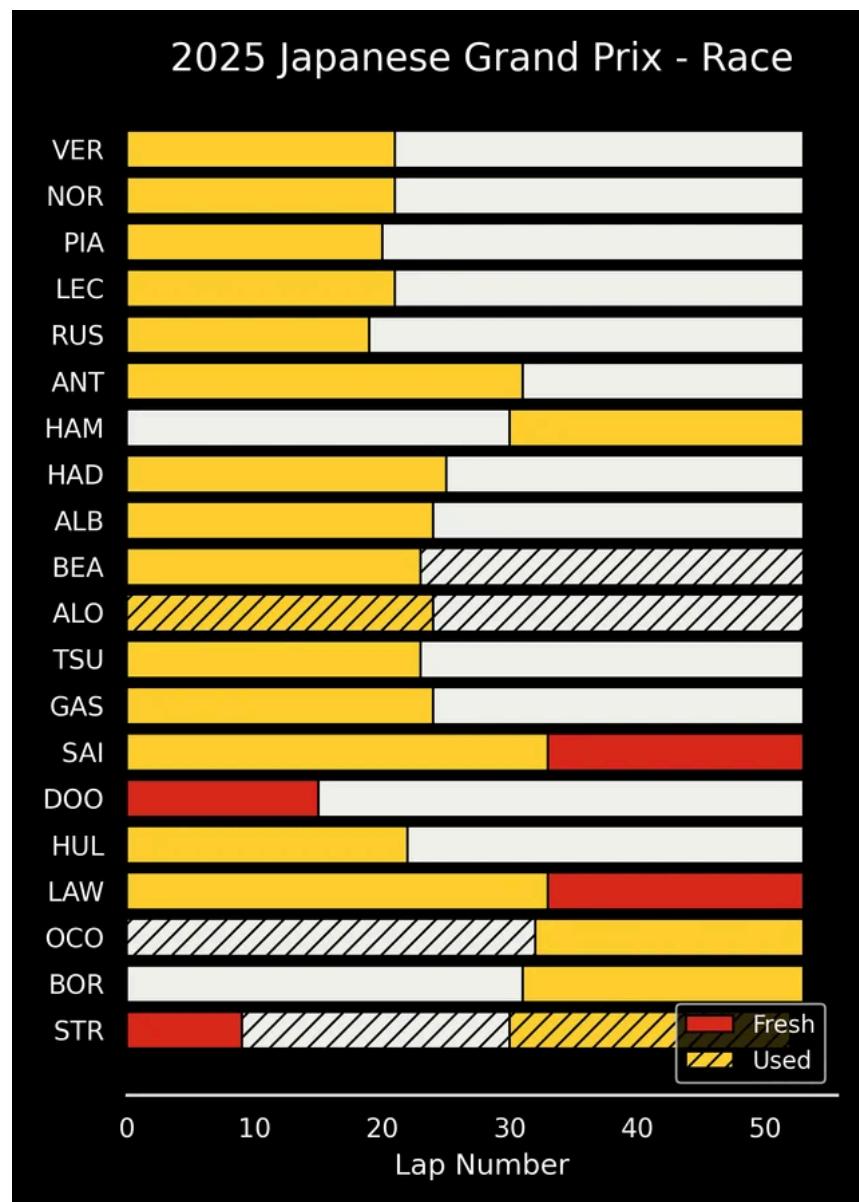
Predviđanje projekta:

1. 1-Stop: SREDNJE (L1-približno L12-18) -> JAKO (L16-56), Vrijeme: 5525.036 s
2. 1-stop: SREDNJE (L1-približno L7-13) -> JAKO (L11-56), vrijeme: 5531.244 s
3. 2-stop: MEKO (L1-približno L5-11) -> SREDNJE (L9-približno L13-19) -> JAKO (L17-56), Vrijeme: 5549.756s

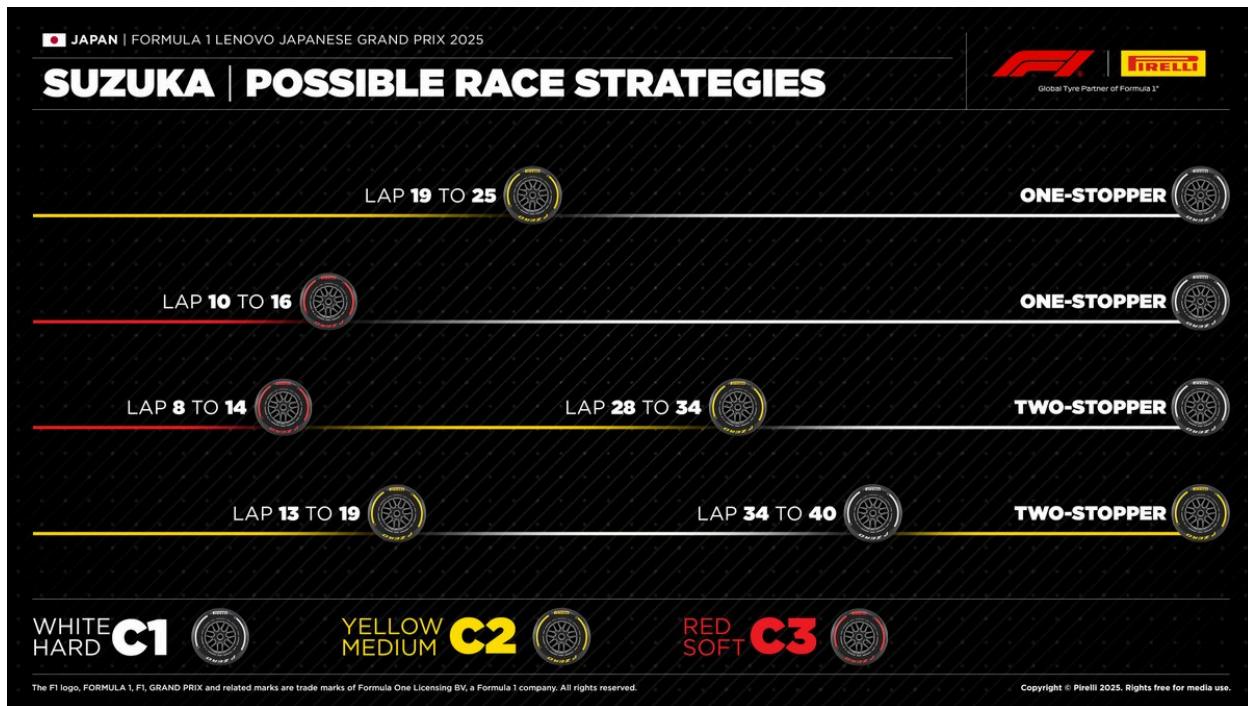
Napomena: ovdje, **naše predviđanje je bilo bolje od Pirellijevog**, jer je većina vodećih vozača bila na 1-stop strategiji od srednjeg do mekog. Ovo je bolji rezultat od očekivanog, jer Kina dugo nije bila u kalendaru trka.

5.1.3. Suzuka International Racing Course (Japan)

- **Karakteristike kola:** Suzuka je legendarni, brzi, tečni krug sa kulnim uglovima poput Esses. To je pravi test aerodinamičke ravnoteže automobila i hrabrosti vozača.
- **Ključni parametri:** Krugovi utrke: 53, prosječno Pit Stop Vrijeme: ~23.5s, Hyp. 2025 Temperatura staze: ~30°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C1), Srednje (C2), Meko (C3).
- **Opći uvid u strategiju:** Budući da je to visokoenergetska staza s najtvrdim spojevima, izdržljivost se često nagrađuje. Strategije sa 1 zaustavljanjem su uobičajene, ali priroda velike brzine i dalje može dovesti do degradacije, čineći 2 zaustavljanja konkurentnim.
- **Rezultati:**



Pirelli predviđanje:



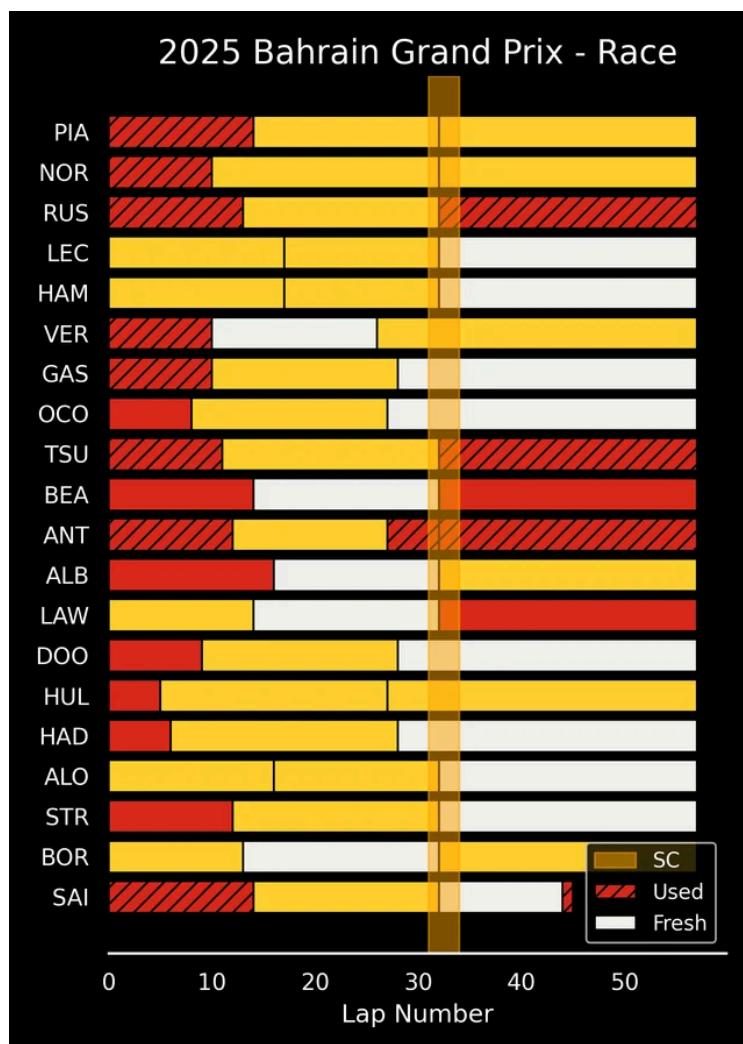
Predviđanje projekta:

1. 1-stop: SREDNJE (L1-približno L22-28) -> JAKO (L26-53), vrijeme: 4976.235 s
2. 1-stop: SREDNJE (L1-približno L17-23) -> JAKO (L21-53), vrijeme: 4990.438 s
3. 2-stop: MEKO (L1-približno L5-11) -> SREDNJE (L9-približno L23-29) -> JAKO (L27-53),
Vrijeme: 5010.515 s

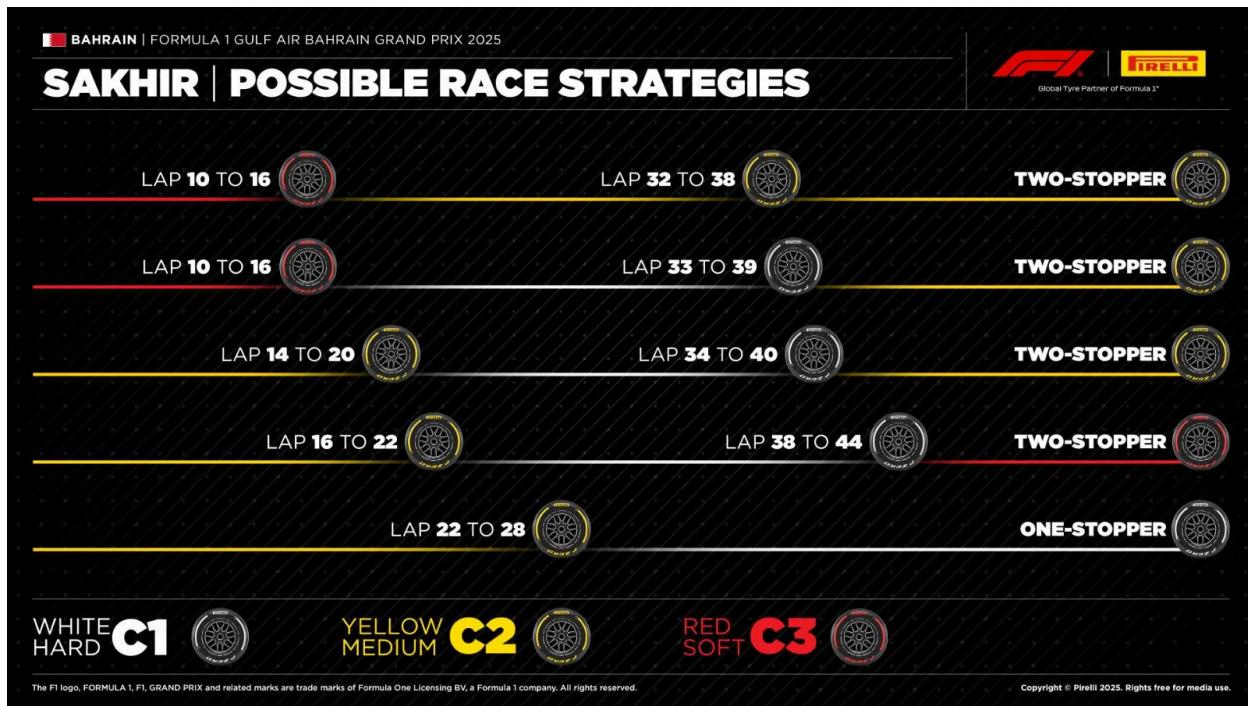
Napomena: odličan rezultat, čak 2 stopera je jako blizu Pirellijevog.

5.1.4. Bahrein International Circuit (Sakhir)

- **Karakteristike kola:** Bahrein je poznat po svojoj abrazivnoj površini staze, izazovnim zonama kočenja i mješavini brzih ravnih i tehničkih krivina. Često se izvodi noću, što može dovesti do nižih temperatura na stazi od dnevnih treninga.
- **Ključni parametri:** Krugovi utrke: 57, prosječno Pit Stop Vrijeme: ~25.0s, Hyp. 2025 Temperatura staze: ~31°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C1), Srednje (C2), Meko (C3).
- **Opći uvid u strategiju:** S obzirom na visoku degradaciju koja se ovdje često vidi, strategija sa 2 zaustavljanja je često optimalna, favorizirajući tvrde smjese za duže staze i mekše smjese za brže sprintove.
- **Rezultati:**



Pirelli predviđanje:



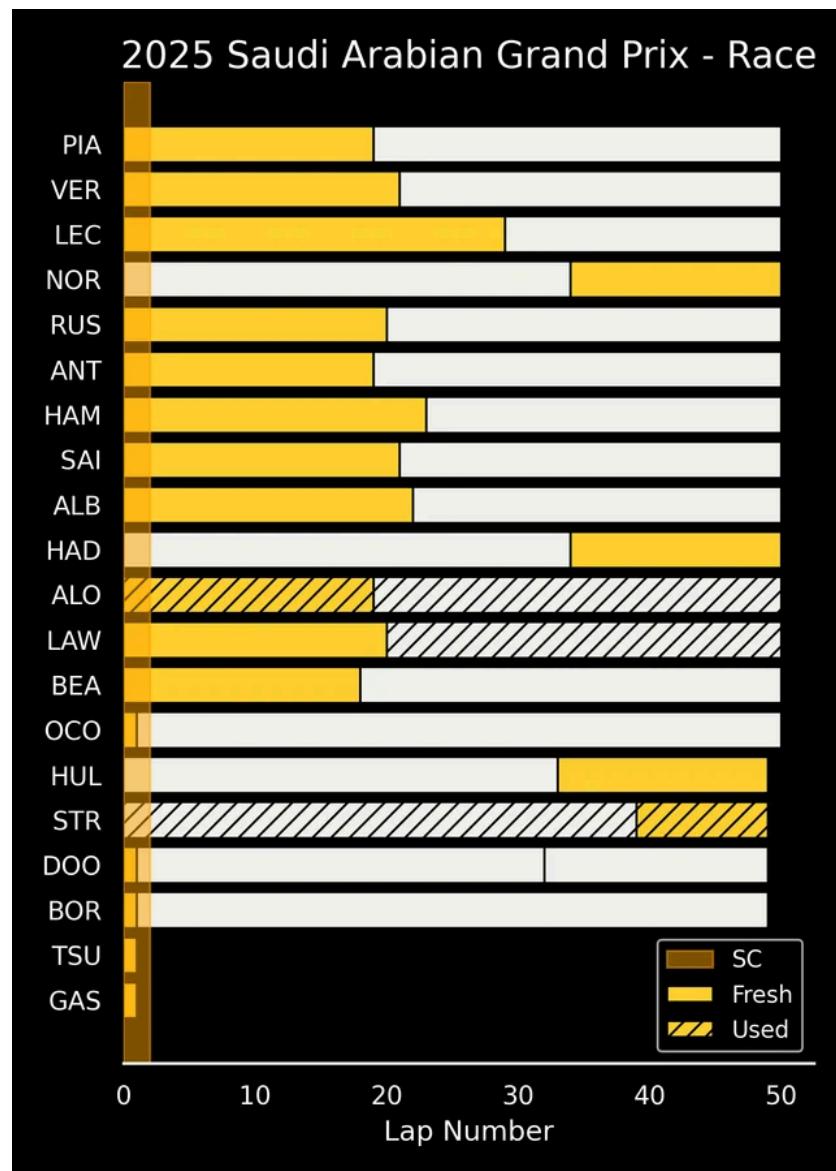
Predviđanje projekta:

1. 1-Stop: MEKO (L1-približno L12-18) -> SREDNJE (L16-57), Vrijeme: 5766.540 s
2. 1-Stop: MEKO (L1-približno L7-13) -> SREDNJE (L11-57), Vrijeme: 5767.337 s
3. 2-stop: MEKO (L1-približno L10-16) -> SREDNJE (L14-približno L43-49) -> MEKO (L47-57), Vrijeme: 5790.097 s

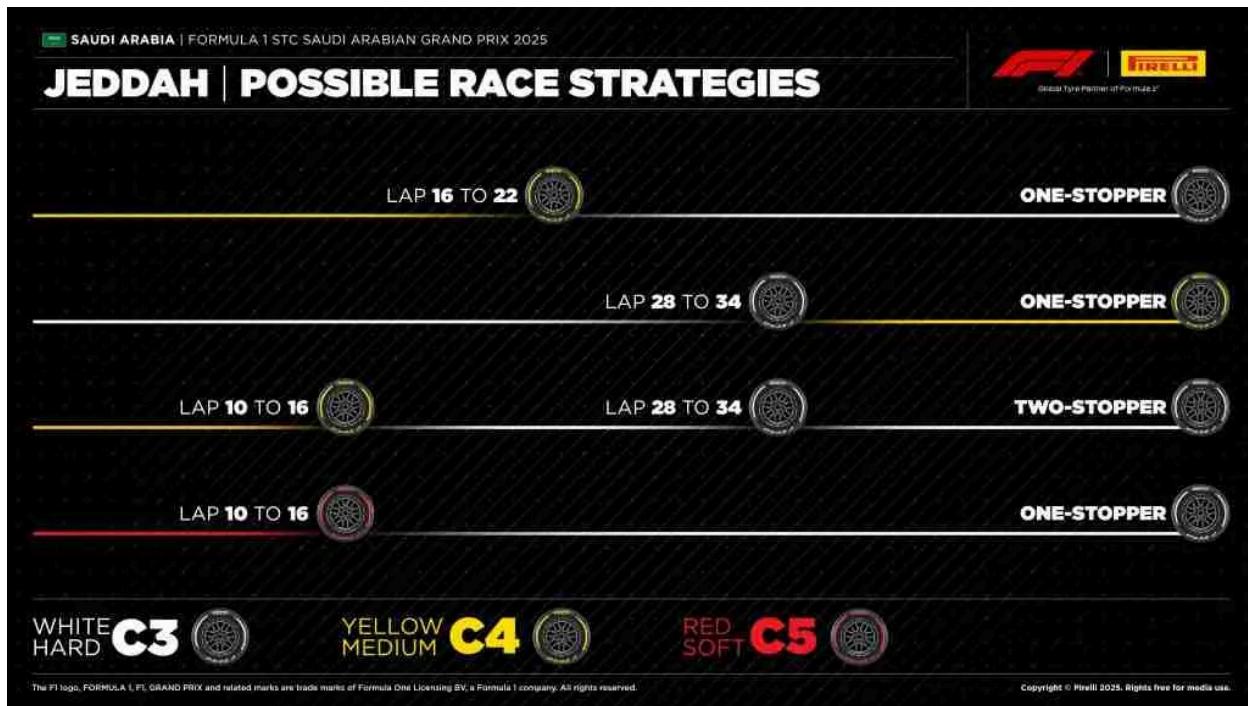
Napomena: ovo je **najslabija tačka ovog projekta**. Neki vozači su zaista koristili gornju strategiju sa 2 zaustavljanja, ali 1 zaustavljanje nije bilo. Međutim, ako pogledamo rezultate utrke iznad, vidimo da je sigurnosni automobil bio na stazi u 32. krugu, što je vozače natjeralo na jeftino zaustavljanje u boksu, što je moguće promijenilo njihovu strategiju sa jednog zaustavljanja na dva zaustavljanja.

5.1.5. Jeddah Corniche Circuit (Saudska Arabija)

- **Karakteristike kola:** Jeddah je ulična staza velike brzine s brojnim brzim krivinama i dugim ravnim dijelovima. Zahteva veliku potisnu silu i pruža ograničene mogućnosti preticanja uprkos svojoj brzini.
- **Ključni parametri:** Krugovi utrke: 50, prosječno Pit Stop Vrijeme: ~19.2s, Hyp. 2025 Temperatura staze: ~36°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C2), Srednje (C3), Meko (C4).
- **Opći uvid u strategiju:** Zbog prirode velike brzine i relativno niske degradacije u poređenju sa Bahreinom, ovdje se često preferira strategija od 1 zaustavljanja ili 2 zaustavljanja ako trošenje guma postane značajan faktor.
- **Rezultati:**



Pirelli predviđanje:



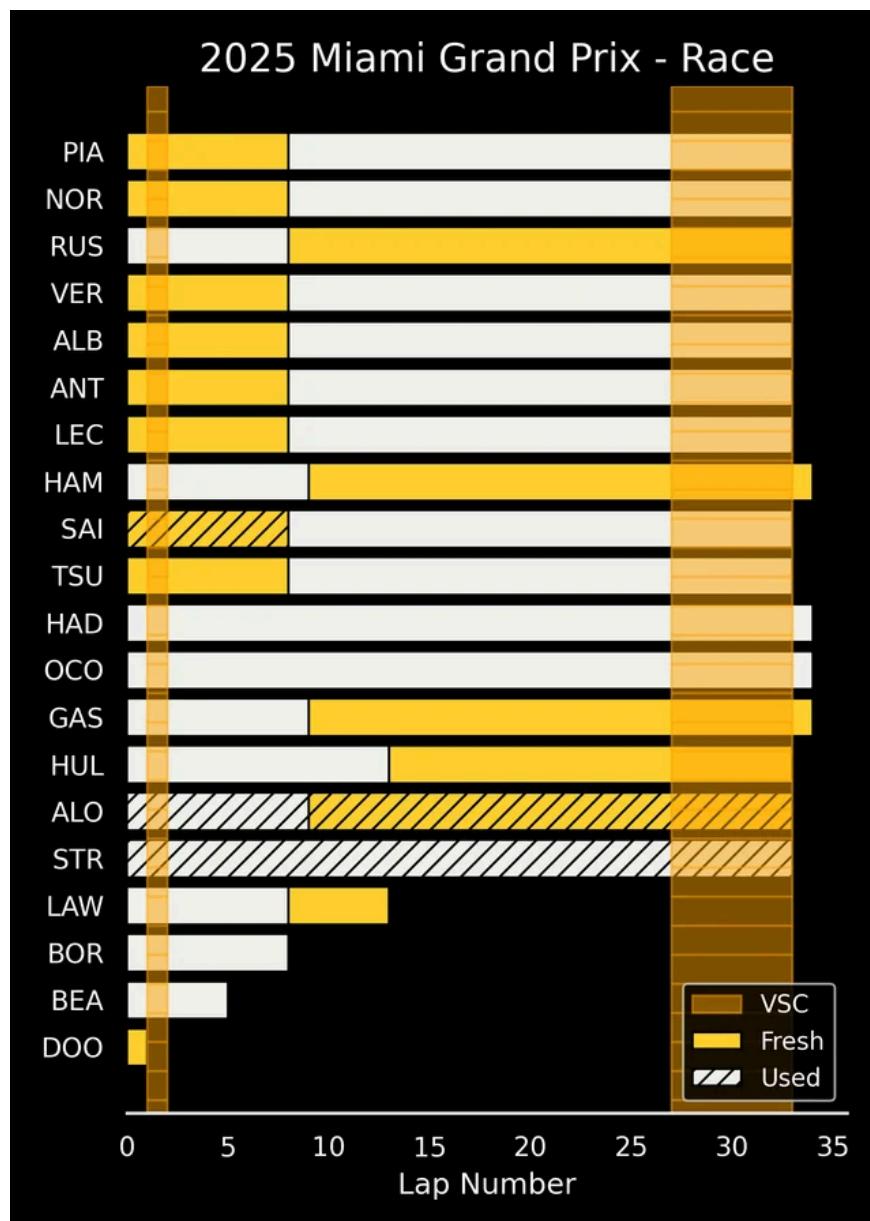
Predviđanje projekta:

1. 1-stop: SREDNJE (L1-približno L22-28) -> TEŠKO (L26-50), vrijeme: 4797.795 s
2. 1-Stop: SREDNJE (L1-približno L17-23) -> JAKO (L21-50), Vrijeme: 4803.668 s
3. 2-stop: MEKO (L1-približno L5-11) -> SREDNJE (L9-približno L23-29) -> JAKO (L27-50),
Vrijeme: 4822.860 s

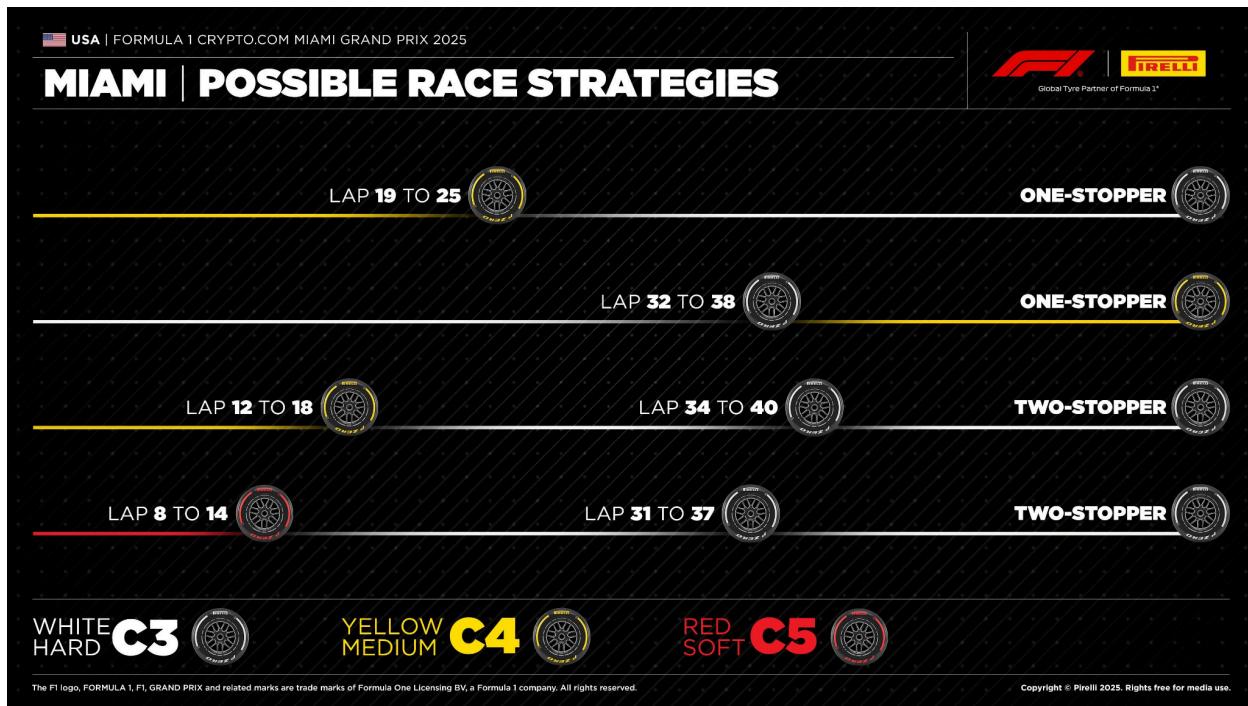
Napomena: Vidimo prve 2 strategije koje koristi skoro svaki vozač, tako da je ovo prilično dobar rezultat.

5.1.6. Miami International Autodrome (SAD)

- **Karakteristike kola:** Namjenski izgrađena ulična staza oko stadiona Hard Rock, koja sadrži mješavinu dionica za velike brzine i spore tehničke šikane. Često je vruće i vlažno, što utiče na performanse guma.
- **Ključni parametri:** Krugovi utrke: 57, prosječno Pit Stop Vrijeme: ~20.0s, Hyp. 2025 Temperatura staze: ~35°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C2), Srednje (C3), Meko (C4).
- **Opći uvid u strategiju:** Uz potencijalno visoke temperature staze, degradacija guma će biti faktor. I 1-stop i 2-stop strategije će vjerovatno biti u sporu.
- **Rezultati:**



Pirelli predviđanje:



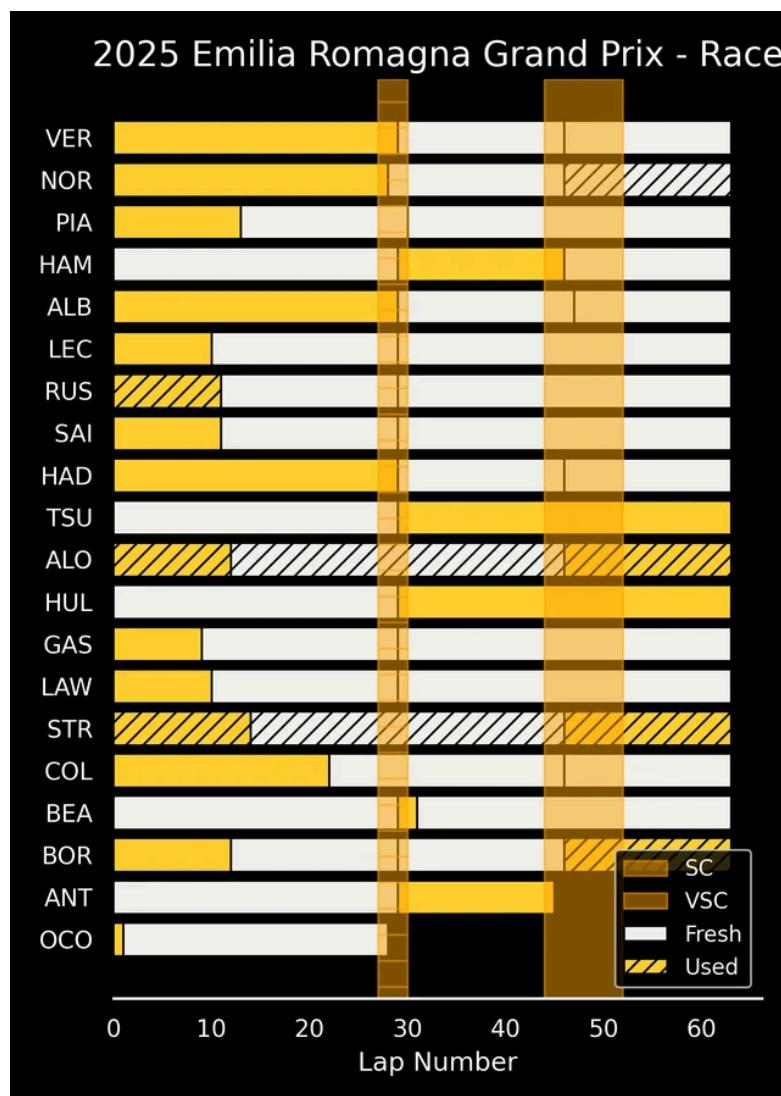
Rezultat projekta:

1. 1-stop: SREDNJE (L1-približno L12-18) -> JAKO (L16-57), vrijeme: 5377.916 s
2. 1-stop: SREDNJE (L1-približno L17-23) -> JAKO (L21-57), vrijeme: 5378.381 s
3. 2-stop: JAKO (L1-približno L25-31) -> SREDNJE (L29-približno L33-39) -> JAKO (L37-57), vrijeme: 5402.098 s

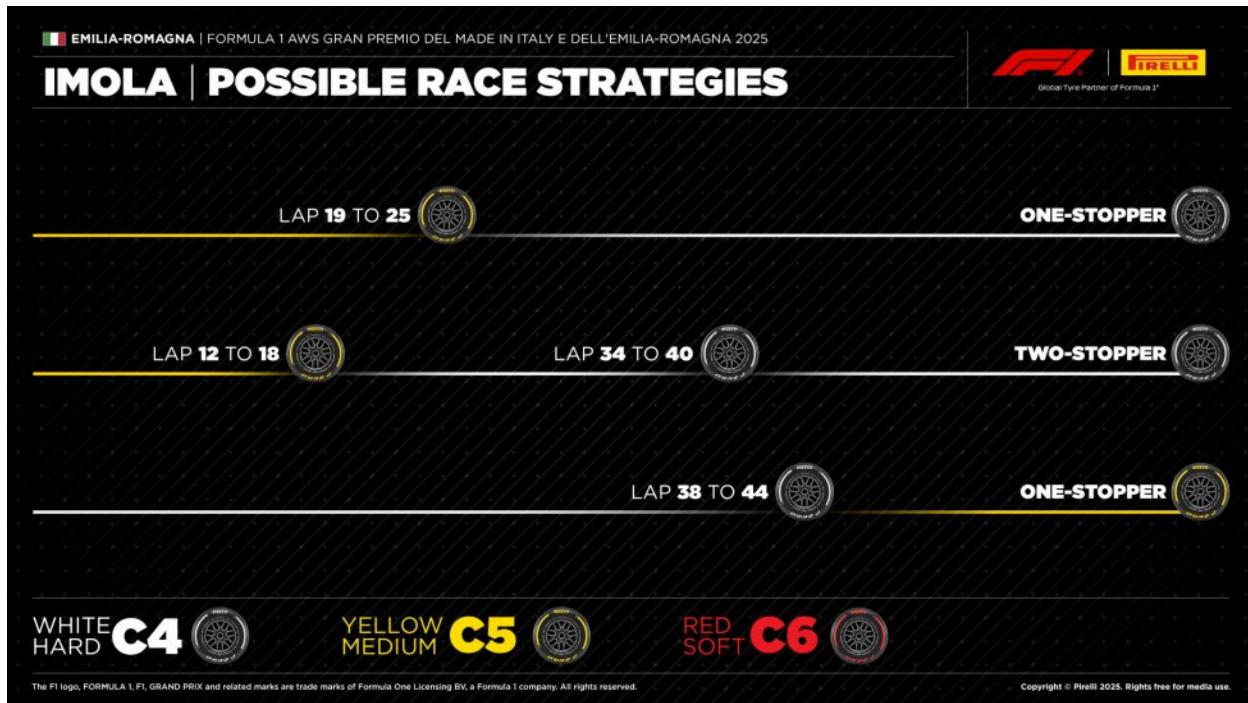
Napomena: Gotovo svi su koristili srednju do maku strategiju, ali su prošli prije 10. kruga.
Međutim, možemo vidjeti da je naša prva strategija zapravo najbliža toj od svih ovih.

5.1.7. Autodromo Internazionale Enzo e Dino Ferrari (Imola, Italija)

- **Karakteristike kola:** Imola je klasična staza stare škole sa skućenim i tehničkim rasporedom, poznata po promjenama nadmorske visine i šikanama. Preticanja je notorno teško.
- **Ključni parametri:** Krugovi utrke: 63, prosječno Pit Stop Vrijeme: ~28.2s, Hyp. 2025 Temperatura staze: ~32°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C3), Srednje (C4), Meko (C5).
- **Opći uvid u strategiju:** S obzirom na poteškoće u preticanju, položaj staze je najvažniji. Ovo često dovodi do toga da timovi daju prioritet strategijama od 1 zaustavljanja kako bi minimizirali vrijeme u pit laneu, uprkos mekšoj dodjeli guma.
- **Rezultati:**



Pirelli predviđanje:



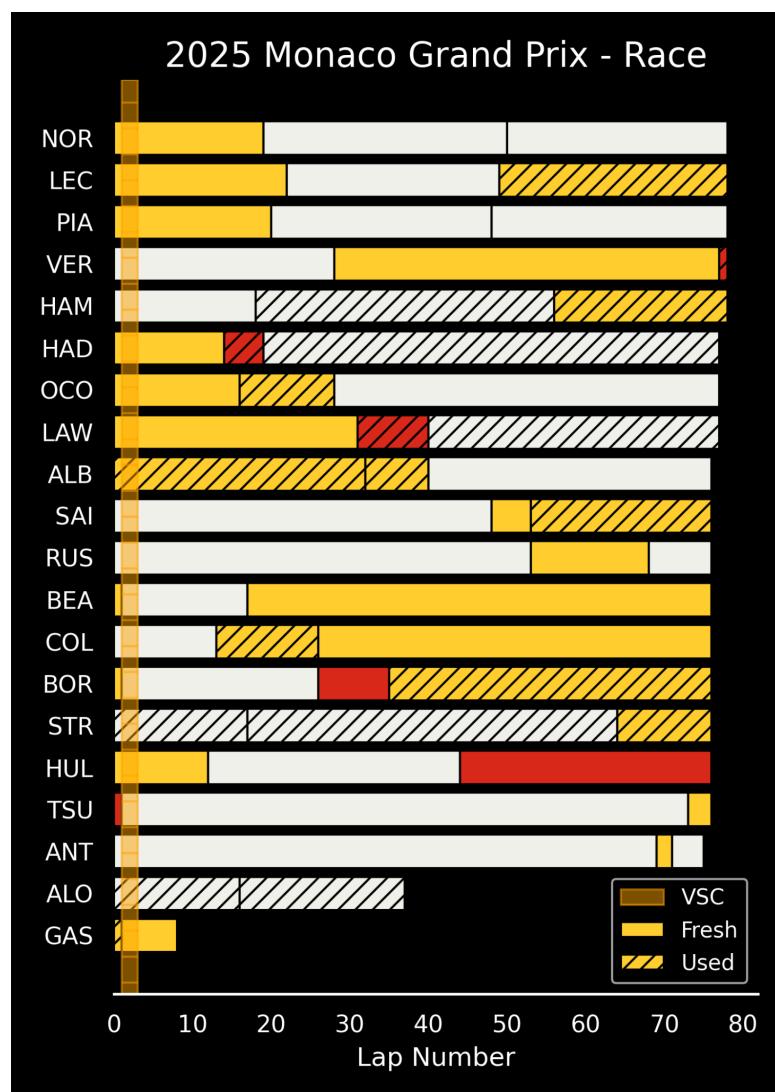
Rezultat projekta:

1. 1-Stop: SREDNJE (L1-približno L47-53) -> MEKO (L51-63), Vrijeme: 5515.617 s
2. 1-Stop: SREDNJE (L1-približno L32-38) -> MEKO (L36-63), Vrijeme: 5527.318 s
3. 2-stop: SREDNJE (L1-približno L30-36) -> MEKO (L34-približno L43-49) -> SREDNJE (L47-63), Vrijeme: 5549.049s

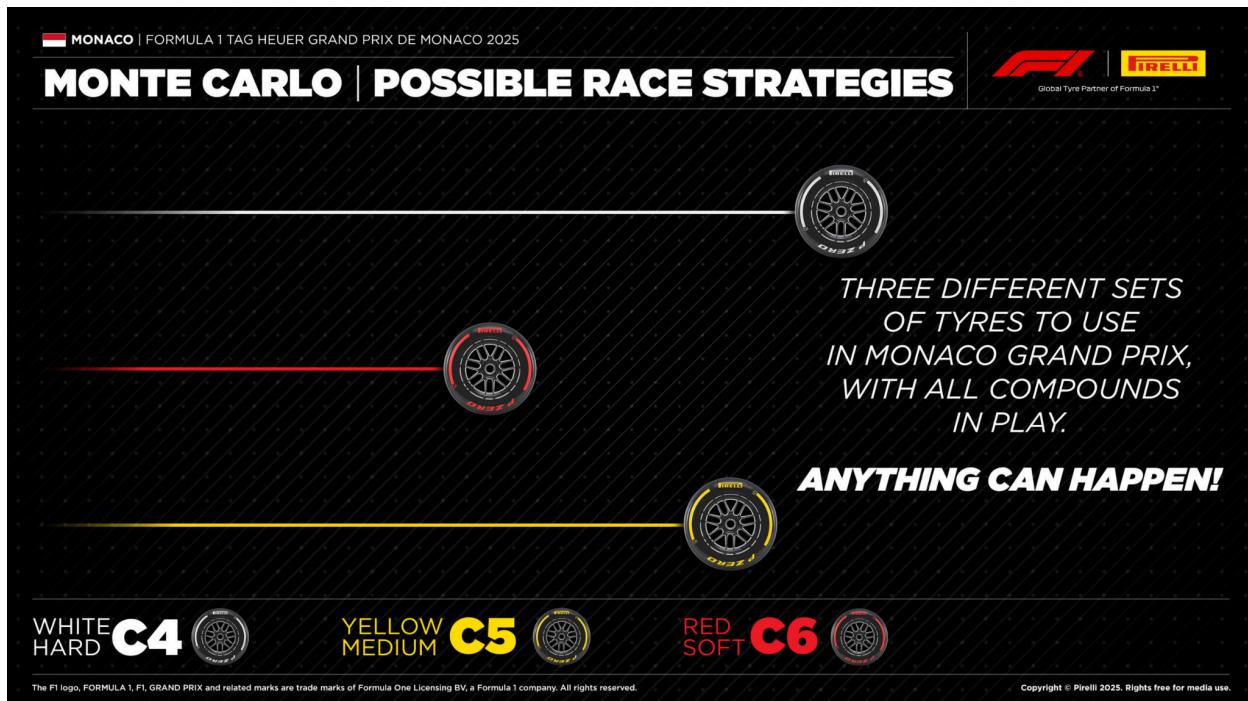
Napomena: slaba tačka naše strategije ovdje je činjenica da je meka guma najmekša guma klase C6. Imaли smo više sigurnosnih automobila u drugom dijelu utrke i to je zaista učinilo meke gume previše rizičnim za odabir.

5.1.8. Circuit de Monaco (Monaco)

- **Karakteristike kola:** Kvintesencijalno ulično kolo. Nevjerovatno čvrsto, sporo i izazovno, sa vrlo malo mesta za greške. Preticanje je gotovo nemoguće.
- **Ključni parametri:** Krugovi utrke: 78, prosječno Pit Stop Vrijeme: ~19.4s, Hyp. 2025 Temperatura staze: ~43°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C3), Srednje (C4), Meko (C5).
- **Opći uvid u strategiju:** Zbog skoro nemogućnosti preticanja i minimalne degradacije guma, Monako je skoro isključivo trka sa 1 zaustavljanjem. Vrijeme zaustavljanja u jednom boksu je kritično, često se poklapa sa sigurnosnim automobilom.
- **Rezultati:**



Pirelli predviđanje:



MONACO | FORMULA 1 TAG HEUER GRAND PRIX DE MONACO 2025

MONTE CARLO | POSSIBLE RACE STRATEGIES

F1 PIRELLI Global Tyre Partner of Formula 1[®]

THREE DIFFERENT SETS OF TYRES TO USE IN MONACO GRAND PRIX, WITH ALL COMPOUNDS IN PLAY.

ANYTHING CAN HAPPEN!

WHITE HARD C4 

YELLOW MEDIUM C5 

RED SOFT C6 

The F1 logo, FORMULA 1, F1, GRAND PRIX and related marks are trade marks of Formula One Licensing BV, a Formula 1 company. All rights reserved.

Copyright © Pirelli 2025. Rights free for media use.

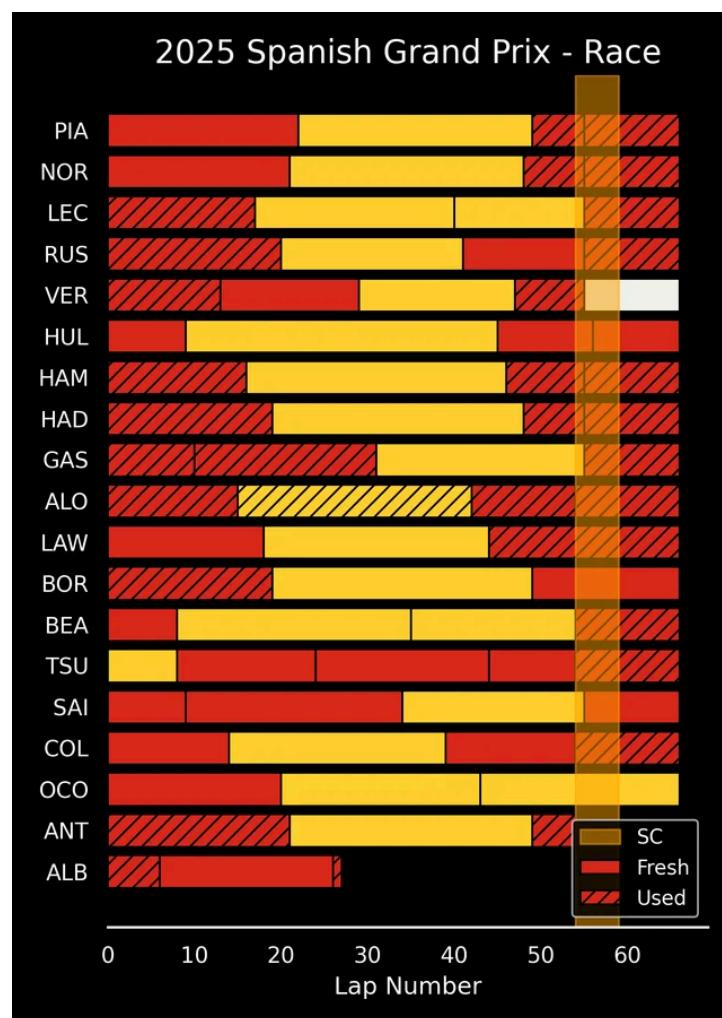
Predviđanje projekta:

1. 1-stop: ČVRSTO (L1-približno L57-63) -> MEKO (L61-78), Vrijeme: 6940.355 s
2. 1-Stop: SOFT (L1-približno L12-18) -> HARD (L16-78), Vrijeme: 6942.570 s
3. 2-stop: MEKO (L1-približno L10-16) -> ČVRSTO (L14-približno L58-64) -> MEKO (L62-78), Vrijeme: 6963.370 s

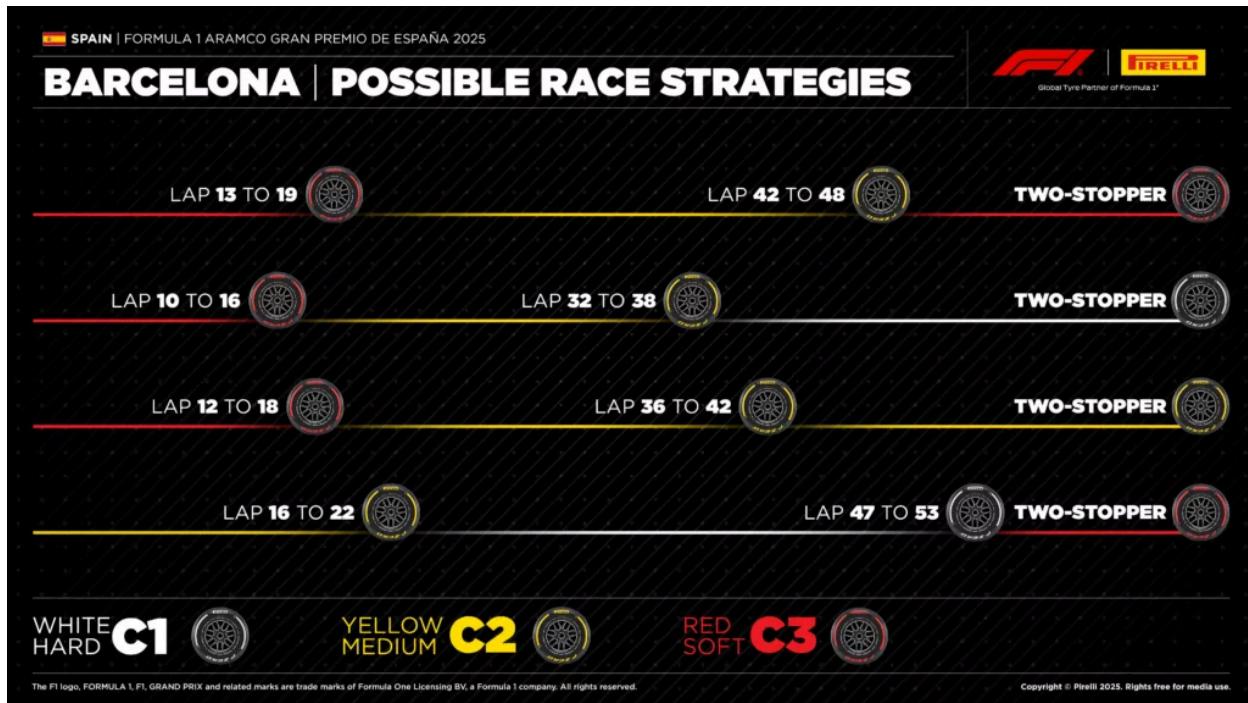
Napomena: Monako je a **SPECIFIČNO** rase. **Novo pravilo uvedeno za 2025.** kaže da vozači moraju koristiti 2 pitstopa za Monako, što zaista onemogućava predviđanje s našim modelom. Pirellijeva strategija je dobro objašnjenje 😊 Međutim, Monako je poznat po tome koliko je teško prestići, tako da su naše strategije prilično dobre (ako bi postojalo pravilo 1 zaustavljanja)

5.1.9. Circuit de Barcelona-Catalunya (Španija)

- **Karakteristike kola:** Dobro zaokružen krug koji se intenzivno koristi za testiranje, pružajući mješavinu brzih krivina, dugih pravaca i tehničkih dionica. Poznat po velikoj degradaciji guma, posebno na prednjoj lijevoj strani.
- **Ključni parametri:** Krugovi utrke: 66, prosječno Pit Stop Vrijeme: ~19.6s, Hyp. 2025 Temperatura staze: ~38°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C1), Srednje (C2), Meko (C3).
- **Opći uvid u strategiju:** Barselonine visoke karakteristike degradacije, čak i sa tvrdim jedinjenjima, često čine trku sa 2 zaustavljanja. Cilj je upravljati habanjem guma uz održavanje snažnog tempa.
- **Rezultati:**



Pirelli predviđanje:



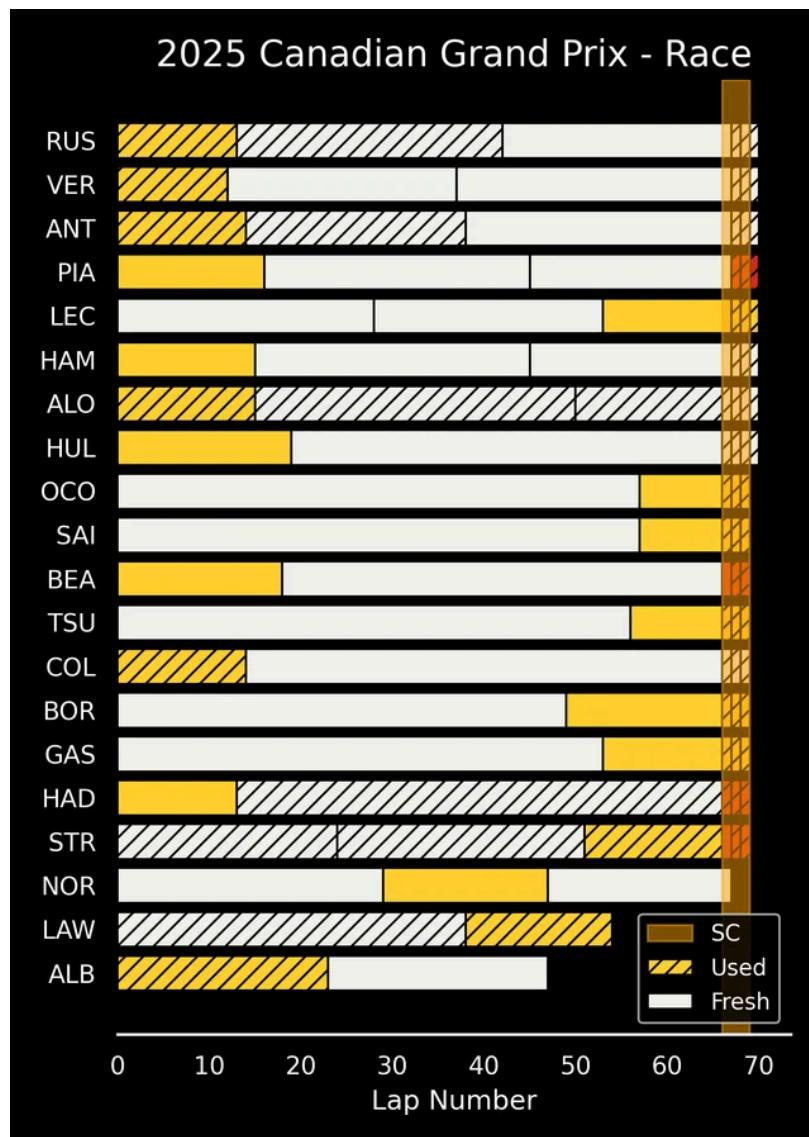
Predviđanje projekta:

1. 1-Stop: MEKO (L1-približno L17-23) -> SREDNJE (L21-66), Vrijeme: 5575.175 s
2. 1-Stop: MEKO (L1-približno L22-28) -> SREDNJE (L26-66), Vrijeme: 5581.346 s
3. 2-stop: MEKO (L1-približno L20-26) -> SREDNJE (L24-približno L48-54) -> MEKO (L52-66), Vrijeme: 5611.859s

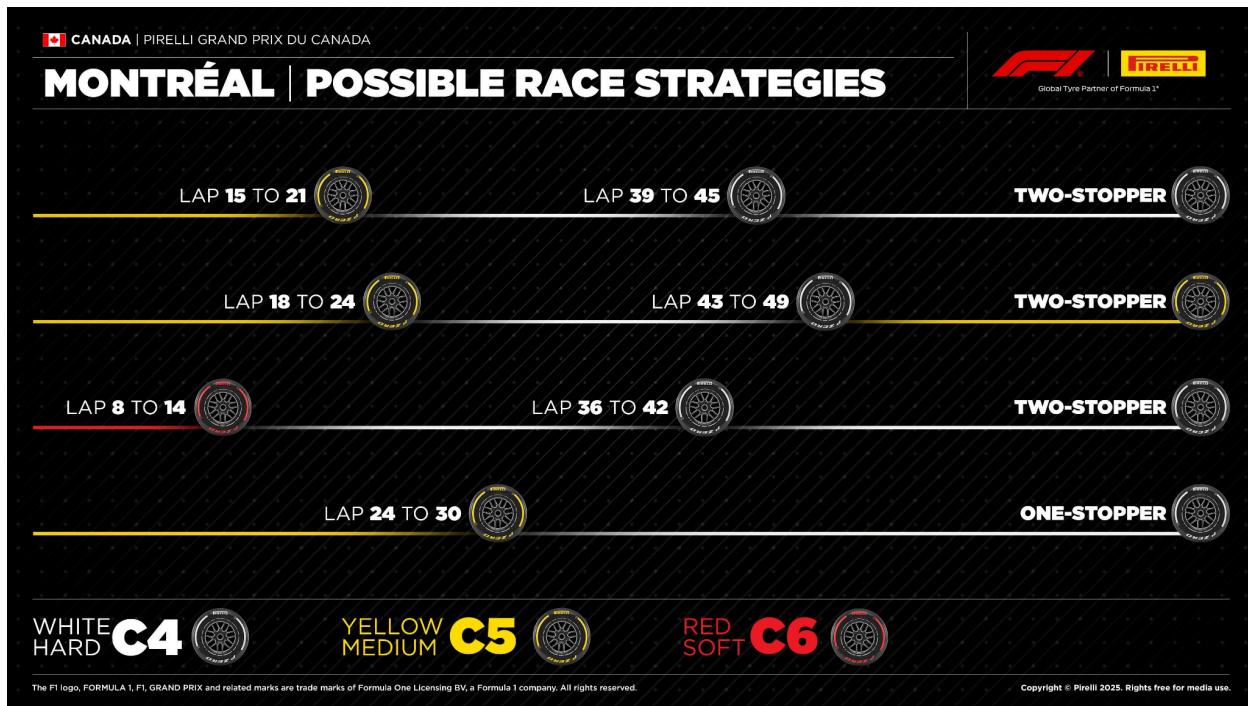
Napomena: Čini se da je naša treća strategija savršena. Ali, razlog zašto se prva 2 ne koriste je taj Barselona staza **doradjen prosle godine**(promijenjena je zadnja krivina, zbog čega je staza koristila mnogo više guma).

5.1.10. Staza Gilles Villeneuve (Kanada)

- **Karakteristike kola:** Jedinstvena polutrajna ulična staza sa dugim ravnima odvojenim šikanama, što dovodi do zona teških kočenja. Poznat po "Zidu šampiona" na izlasku iz finalne šikane.
- **Ključni parametri:** Krugovi utrke: 70, prosječno Pit Stop Vrijeme: ~18.4s, Hyp. 2025 Temperatura staze: ~30°C.
- **Predviđena alokacija kompleksa za 2025.** Tvrdo (C3), Srednje (C4), Meko (C5).
- **Opći uvid u strategiju:** Uz značajno kočenje i ubrzanje, degradacija guma može biti velika. Često se viđaju i strategije s 1 i 2 zaustavljanja, ovisno o tome kako timovi upravljaju mekšim spojevima.
- **Rezultati:**



Pirelli predviđanje:



Rezultat projekta:

1. 1-stop: SREDNJE (L1-približno L7-13) -> JAKO (L11-70), vrijeme: 5567.644 s
2. 1-Stop: SREDNJE (L1-približno L12-18) -> JAKO (L16-70), Vrijeme: 5580.930 s
3. 2-stop: JAKO (L1-približno L25-31) -> SREDNJE (L29-približno L43-49) -> JAKO (L47-70), vrijeme: 5583.430 s

Napomena: Prve 2 strategije koriste mnogi vozači, međutim, neki koriste 2 stope SREDNJE-TVRDO->TEŠKO. Razlog tome je što se na treninzima koristilo mnogo tvrdih guma, pa su vozači morali 2 puta u boks.

5.2. Poređenje sa radovima iz prethodne faze

Ovaj projekat doprinosi postojećem stanju tehnike u primeni ML-a na sportske analize, posebno u Formuli 1. Međutim, u poređenju sa najnaprednijim komercijalnim alatima koje koriste F1 timovi, ovaj projekat je na osnovnom nivou. Komercijalni sistemi često uključuju podatke u realnom vremenu, složenije fizičke modele guma i automobila, detaljno modeliranje saobraćaja, predviđanje ponašanja takmičara i sposobnost prilagođavanja strategije u deliću sekunde tokom trke.

5.3. Diskusija o mogućim poboljšanjima

Iako su postignuti značajni rezultati, uvijek ima prostora za napredak:

- **Dinamički vremenski/uvjeti na stazi:** Trenutna simulacija koristi fiksnu hipotetičku temperaturu i nema kiše. Integracija vremenske prognoze u realnom vremenu, ili čak vjerovatnostnih modela promjena stanja staze tokom trke, značajno bi poboljšala realizam.
- **Napredno modeliranje guma:** Implementacija eksplicitnih krivulja degradacije pneumatika za svaku smjesu i stazu, koje bi se ažurirale u realnom vremenu ili predviđale na osnovu uslova na stazi i agresivnosti vozača.
- **Adaptacija strategije vođena događajima:** Uvođenje logike za automatsku reakciju na nepredviđene događaje kao što su sigurnosni automobili, virtualni sigurnosni automobili, crvene zastavice ili neočekivani problemi u pit lane-u.
- **Preporuka višeciljne strategije:** Umjesto samo minimiziranja vremena utrke, razmatranja minimiziranja habanja guma ili optimizacije položaja staze, korištenjem algoritama višestruke optimizacije za predstavljanje raznovrsnijeg skupa preporuka.