

Predykowanie wyników profesjonalnych meczy Dota 2

Aleksandr Shestakov
nr index 272657
Politechnika Wrocławska
Wydział W4N

06 June 2024

Spis treści

1	Cele projektu	2
2	Dane	2
2.1	Cechy	2
2.1.1	Matches diff	3
2.1.2	Winrates diff	3
2.1.3	Counter picks winrate diff	3
2.1.4	Synergy winrate diff	3
2.1.5	Networths diff	4
2.1.6	Signatures diff	4
2.2	Pobieranie danych	4
2.3	EDA	5
2.4	Analiza cech	5
3	Trenowanie modeli	7
4	Wnioski	10

1 Cele projektu

Celem projektu jest badanie zależności wyniku profesjonalnych meczy gry komputerowej Dota 2 od wybranych cech.

2 Dane

2.1 Cechy

Do analizy zostały wybrane następujące cechy:

- Liczba zagranych meczy dla każdego gracza (**Matches**)
- Stosunek wygranych meczy do wszystkich meczy dla każdego gracza (**winrate**)
- Winrate każdego bohatera w każdym zespole w grze przeciwko każdemu bohaterowi w przeciwnym zespole (**Counter picks winrate**)
- Winrate każdego bohatera w każdym zespole w grze z każdym bohaterem w swoim zespole (**Synergies winrate**)
- Ilość złota u każdego gracza na 10 minut gry (**Networths**)
- Czy gracz gra jednym ze swoich najczęściej wybieranych bohaterów (**Signatures**)
- Rankingi graczy w odpowiednich dywizjach (**Ranks**)

Ze względu na liczbę graczy w każdym zespole, wszystkie dane oprócz rankingów zostały opracowane w pojedynczych kolumnach, gdzie znak wartości w tych kolumnach określa która drużyna miała przewagę w wybranej cechce - "+" - Radiant, "-" - Dire:

2.1.1 Matches diff

$$\text{Matches diff} = \sum_{i=1}^5 a_i - \sum_{j=1}^5 b_j$$

Gdzie:

- a_i liczba meczy zagrananych przez gracza i zespołu **Radiant**
- b_j liczba meczy zagrananych przez gracza j zespołu **Dire**

2.1.2 Winrates diff

$$\text{Winrates diff} = \overline{W}_{\text{Radiant}} - \overline{W}_{\text{Dire}}$$

Gdzie:

- $\overline{W}_{\text{Radiant}}$ - średnia winrate'ów graczy zespołu **Radiant**
- $\overline{W}_{\text{Dire}}$ - średnia winrate'ów graczy zespołu **Dire**

2.1.3 Counter picks winrate diff

$$\text{Counter picks winrate diff} = R_{\text{avg}} - D_{\text{avg}}$$

Gdzie:

$$R_{\text{avg}} = \frac{1}{5 \cdot 5} \sum_{i=1}^5 \sum_{j=1}^5 \text{WR}(R_i, D_j)$$

średni Winrate bohaterów w zespole Radiant przeciw bohaterów w Dire, a

$$D_{\text{avg}} = \frac{1}{5 \cdot 5} \sum_{i=1}^5 \sum_{j=1}^5 \text{WR}(D_i, R_j)$$

średni Winrate bohaterów w zespole Dire przeciw bohaterów w Radiant.

Gdzie:

- R_i and D_j oznaczają bohaterów w odpowiednich zespólach.
- $\text{WR}(R_i, D_j)$ oznacza Winrate R_i przeciw D_j .

2.1.4 Synergy winrate diff

$$\text{Synergy winrate diff} = R_{\text{avg}} - D_{\text{avg}}$$

Gdzie:

$$R_{\text{avg}} = \frac{1}{20} \sum_{i=1}^5 \sum_{\substack{j=1 \\ j \neq i}}^5 \text{WR}(R_i, R_j)$$

jest średnią sum Winrateów każdego bohatera z każdym innym w swojej drużynie Radiant

$$D_{\text{avg}} = \frac{1}{20} \sum_{i=1}^5 \sum_{\substack{j=1 \\ j \neq i}}^5 \text{WR}(D_i, D_j)$$

jest średnią sum Winrateów każdego bohatera z każdym innym w swojej drużynie Dire

Tutaj:

- R_i i D_j reprezentują bohaterów w drużynach Radiant i Dire odpowiednio.
- $WR(R_i, R_j)$ oznacza Winrate R_i z R_j w jednym zespole.
- $WR(D_i, D_j)$ oznacza Winrate D_i z D_j w jednym zespole.

2.1.5 Networthes diff

$$\text{Networthes diff} = \overline{N}_{\text{Radiant}} - \overline{N}_{\text{Dire}}$$

Gdzie:

- $\overline{N}_{\text{Radiant}}$ - średnia wartość networth'ów graczy zespołu **Radiant**
- $\overline{N}_{\text{Dire}}$ - średnia wartość networth'ów graczy zespołu **Dire**

2.1.6 Signatures diff

$$\text{Signatures diff} = \sum_{i=1}^5 N s_i - \sum_{j=1}^5 N s_j$$

Gdzie:

- s_i przyjmuje wartość 1 jeżeli gracz i zespołu **Radiant** gra na swojej signature, 0 jeśli nie
- s_j przyjmuje wartość 1 jeżeli gracz j zespołu **Dire** gra na swojej signature, 0 jeśli nie

2.2 Pobieranie danych

Ze względu na brak jednego źródła gdzie znalazłbym wszystkie wymagane dane, pobierałem/skrapowałem ich z kolejnych źródeł:

- **Stratz** - pobierałem za pomocą API tej strony kolejne dane:
 - Imiona wszystkich bohaterów w grze Dota 2
 - Networth graczy w 10 minucie gry
 - Liczba meczy graczy
 - Winrate graczy
- **DotaBuff** - zeskrapowałem kolejne dane:
 - Rankingi graczy, uzupełniałem na 5001 jeśli gracz nie znajdował się w rankingach
 - Signatures graczy
 - Winrate'y wszystkich bohaterów przeciw wszystkim innym
- **Datdota** - bezpośrednio pobrałem csv plik zawierający ID ostatnich 2500 meczy profesjonalnych, ich wynik i inne niewykorzystane informacje
- **Kaggle** - bezpośrednio pobrałem csv plik zawierający dane o winrateach niektórych bohaterów w parze z innymi, dane niedostające uzupełniałem liczbami 0.5

2.3 EDA

Szczegółowy EDA wraz z profilem danych został utworzony w Jupyter Notebook projektu, tutaj pokaże tylko jak wyglądają dane w rzeczywistości:

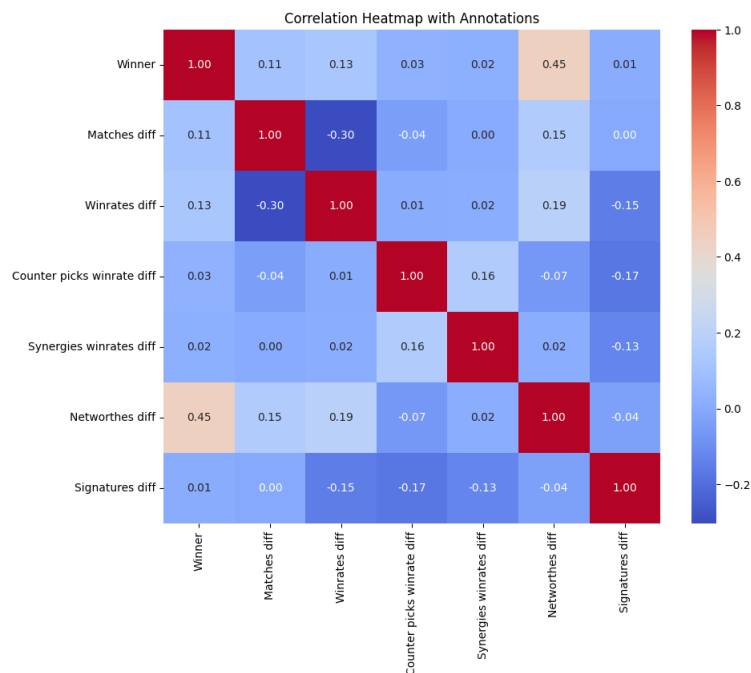
	Winner	Matches diff	Winrates diff	Counter picks winrate diff	Synergies winrates diff	Networthes diff	Signatures diff	R1 rank	R2 rank	R3 rank	R4 rank	R5 rank	D1 rank	D2 rank	D3 rank	D4 rank	D5 rank
0	0	-11333	-0.0084	0.03513	-0.00513	-1918	-1	141	156	15	15	100	19	29	1	6	153
1	0	8208	-0.0116	-0.01034	-0.01126	-1631	0	62	67	3	213	42	19	29	1	6	153
2	0	-8208	0.0116	0.04882	0.02032	3319	-1	19	29	1	6	153	62	67	213	3	42
3	0	8208	-0.0116	-0.06449	-0.00254	1815	1	62	67	213	3	42	19	29	1	6	153
4	1	-7181	0.0064	-0.00085	0.02585	1605	1	19	5	96	37	125	141	156	15	15	100
...
2228	1	4252	0.0044	0.01940	-0.01633	4582	-1	27	122	39	12	255	52	63	66	16	158
2229	1	-4252	-0.0044	-0.00915	-0.01118	0	0	52	63	66	16	158	255	27	39	122	12
2230	1	-21646	0.0086	0.07434	0.00171	2294	-1	313	259	247	259	150	128	393	315	76	78
2231	0	-21646	0.0086	-0.00567	0.03282	526	0	313	259	247	259	150	128	393	315	76	78
2232	1	-21646	0.0086	-0.01288	-0.03754	1771	0	313	247	259	259	150	128	393	315	76	78

2233 rows x 17 columns

Rysunek 1: Tabela danych przed skalowaniem

2.4 Analiza cech

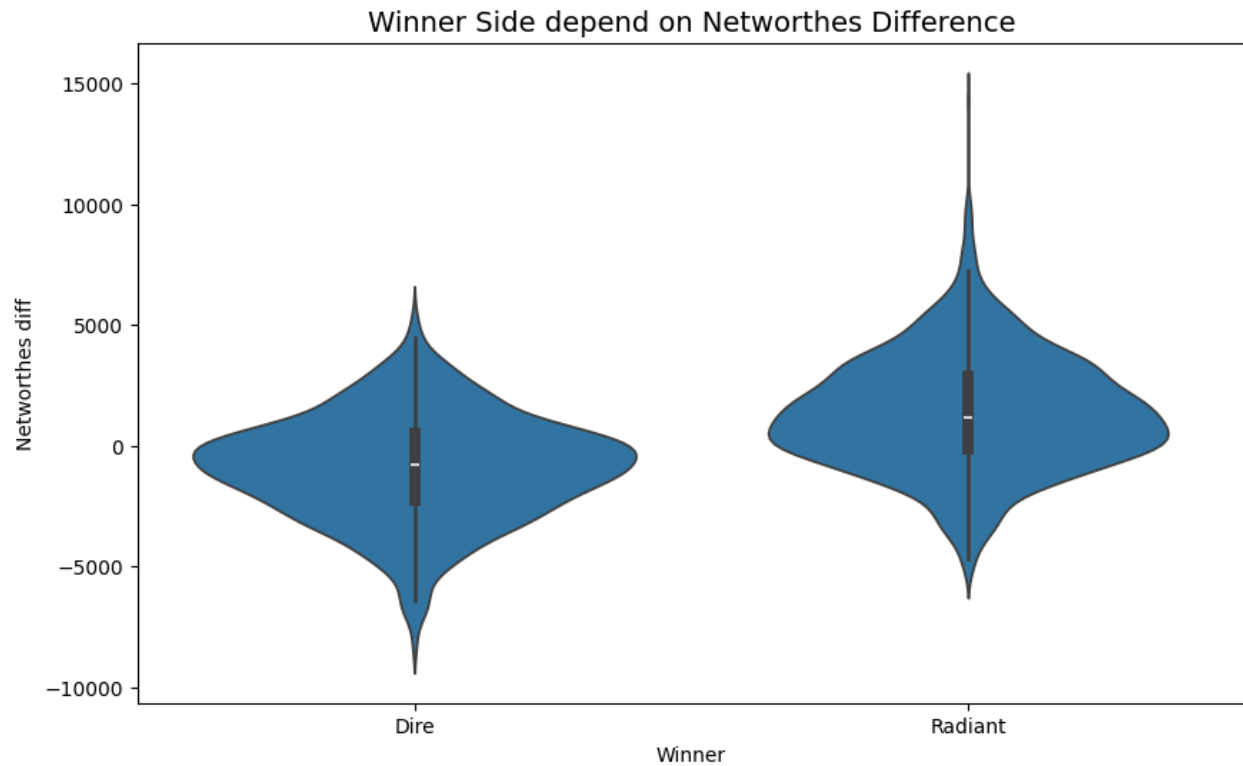
Po przygotowaniu wszystkich danych utworzymy wykres ze wszystkimi korelacjami, z pominięciem rankingów, ponieważ rank pojedynczych graczy nie nadaje się do naszej analizy, i ich udział będziemy rozpatrywać już na etapie trenowania modeli:



Rysunek 2: Heatmapa korelacji

Z heatmapy widzimy że największą korelację z Winner mają **Networthes diff**, **Winrates diff** i **Matches diff**.

Rysowanie dowolnych innych typów wykresów jest bez sensu ze względu na niski poziom korelacji między danymi i ten fakt że rozwiązywany problem jest problemem klasyfikacji(co oznacza że przykładowe violin plots nie nadają się do analizy "na oko" dla wszystkich cech oprócz Networthes diff, gdzie korelacja jest dość wysoka):



Rysunek 3: Zwycięska drużyna na podstawie Networthes diff

Z wykresu widzimy że we większości gier Radiant wygrywa gdy ma przewagę w Networth nad Dire, analogicznie dla Dire.

3 Trenowanie modeli

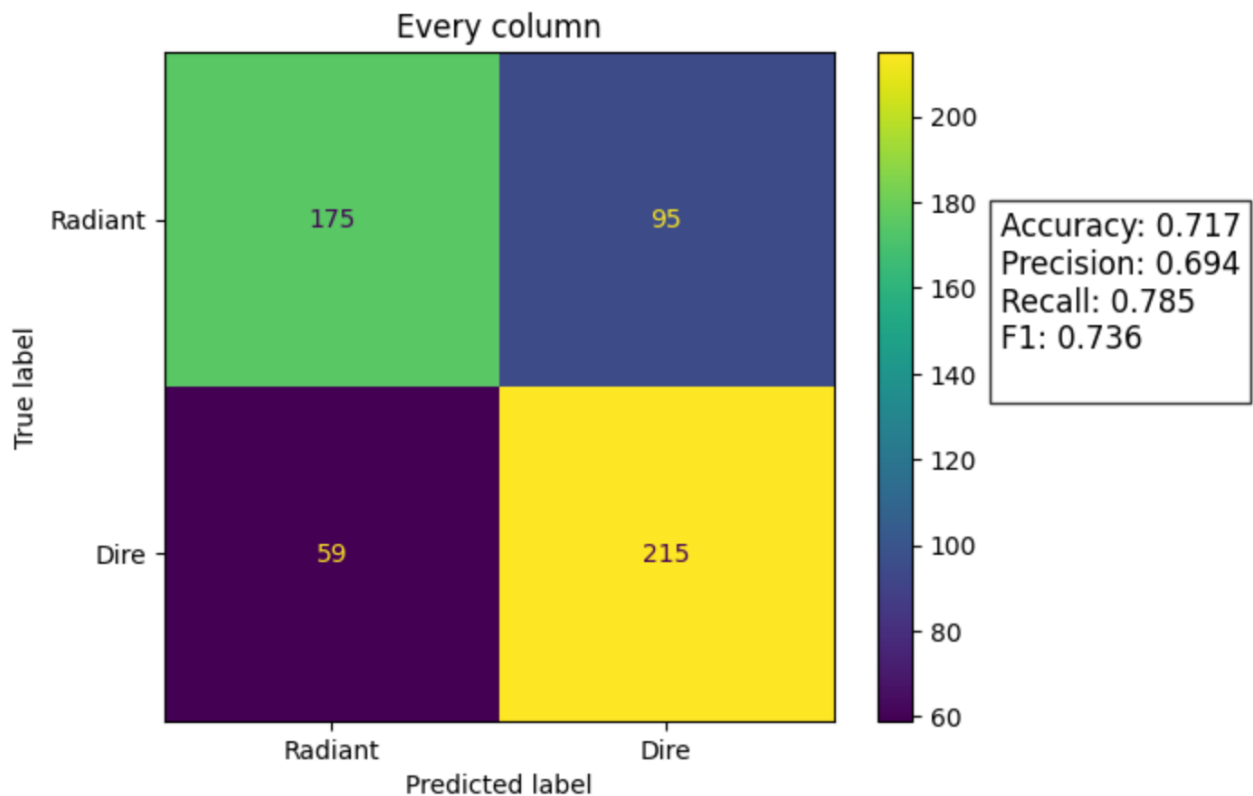
Chociaż z analizy cech wyznaczyliśmy że tylko 3 z nich mają relatywnie dobrą korelację z tym jaka drużyna wygra grę, w celach eksperymentu przetrenujemy kilka modeli na podstawie różnych zbiorów cech.

Trenujemy model na podstawie 75% wszystkich danych.

Zaczynając od trenowania SVM na podstawie wszystkich cech, otrzymujemy najlepsze wyniki dla kolejnych parametrów SVM:

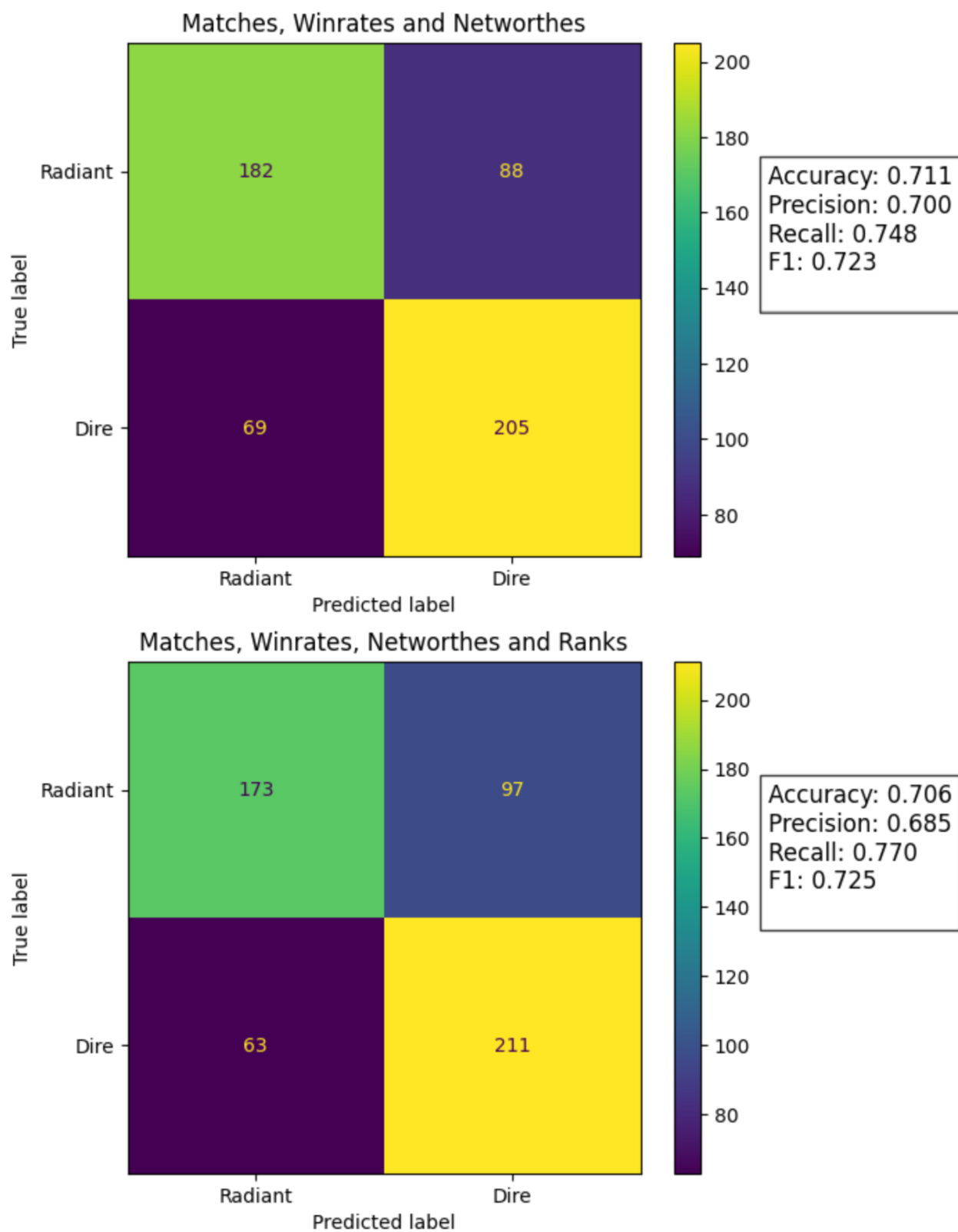
$$\gamma = 0.009 \quad (1)$$

$$C = 231 \quad (2)$$



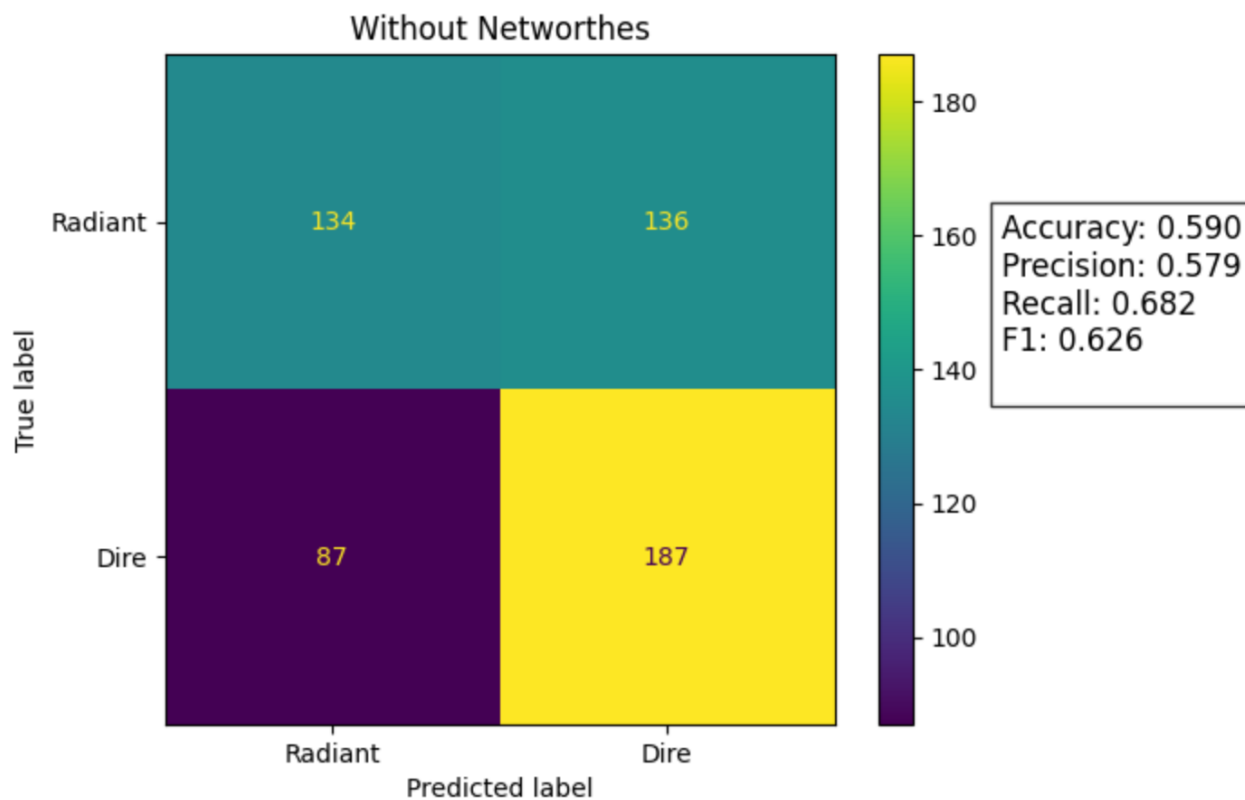
Rysunek 4: Macierz pomyłek i metryki SVM na podstawie wszystkich danych

Jak widać model dość dobrze radzi z przewidywaniem wyniku meczy, sprawdzimy czy nie da się zwiększyć go efektywności przez usunięcie wybranych cech:



Rysunek 5: Widzimy że **rankingi** mają mały wpływ na wynik gry, ale jest on na poziomie odchyłka

Teraza sprawdzimy jak dobrze model poradzi z predykowaniem wyniku gry przed jej rozpoczęciem, tzn. bez danych o **Networth**:



Jak widać, bez danych o **Networth** nie da się w dobry sposób przewidzieć wyniki wszystkich gier. Co oznacza że jedynym sposobem na przewidzenie wyniku gry przed jej początkiem jest przeprowadzenie bardziej skomplikowanej analizy **Bohaterów**, obecnej **Mety** (ang. Most Effective Tactic Available) w grze i samych **graczy**.

4 Wnioski

W tej pracy zbadałem zależność wyniku meczy Dota 2 od podanych cech, i dotarłem do następujących wniosków:

- Na 10 minucie gry da się z dość dobrym *prawdopodobieństwem* $\approx 69\%$ przewidzieć która drużyna wygra mecz, wynika to z tego że dość dużo w meczu zależy od stanu drużyn po stadii **Lining'u**, która kończy się około 10 minuty.
- Przed początkiem gry, jej wynik da się przewidzieć tylko z *prawdopodobieństwem* $\approx 59\%$, wynika to z tego, że wynik meczu najbardziej zależy od dość subiektywny cech, takich jak:
 - Networth'y drużyn
 - Styl gry każdego gracza
 - Obecna Meta w grze
 - Synergie i Counter picks bohaterów, które nie da się wyrazić za pomocą zwykłych liczb
 - I wiele innych...

Z tego wynika ogólna możliwość trenowania o wiele bardziej skomplikowanych modeli, na podstawie literalnie przepływu całych meczy, gdzie model będzie się starać wyznaczyć style gry pojedynczych graczy, uczyć się temu jak bohaterowie odgrywają swoją rolę w meczu i temu które bohaterowie synergują z którymi, a które kontrataktują których.