

Leonardo Baraldi, Riccardo Barbieri, Leonardo Ruberto, Gabriele Tassinari

Requisiti del servizio

Requisiti legati al **feature vector** ...

... funzionalità extra legate all'applicazione web

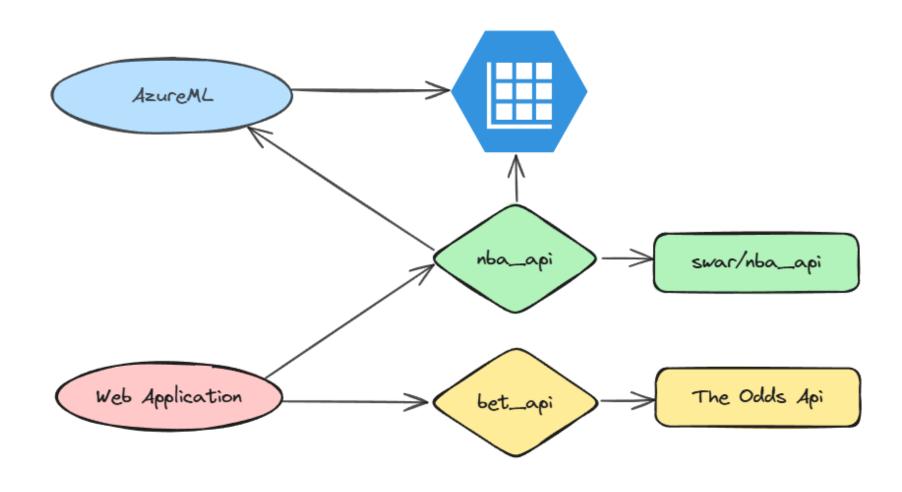
 Informazioni statistiche su partite, squadre e giocatori



• **Quote** di scommesse reali aggiornata in tempo reale



Architettura



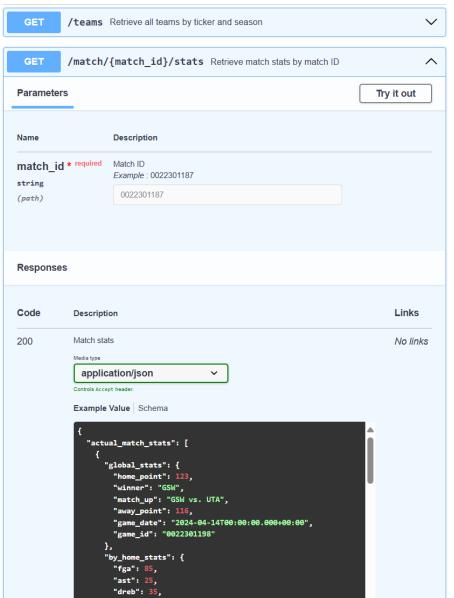
Middleware

REST API Server

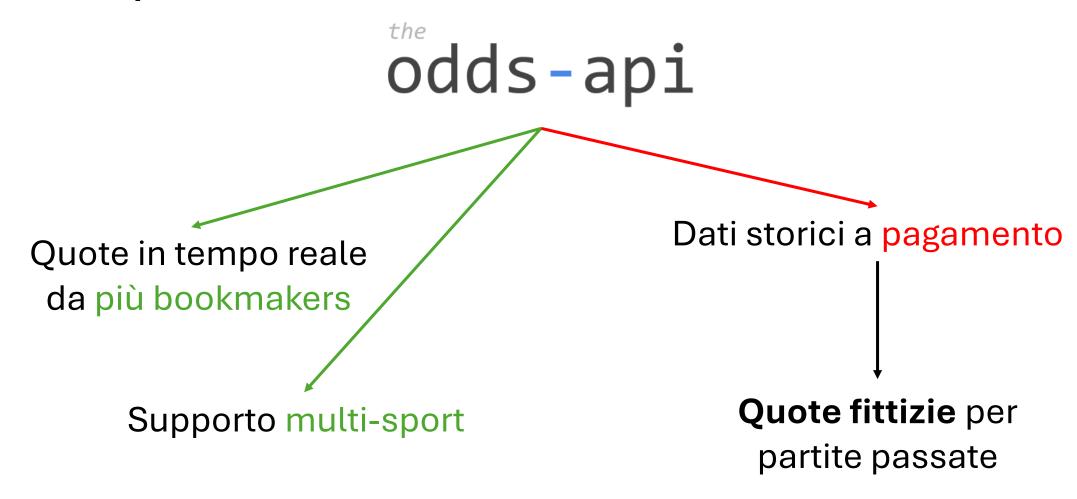
```
openapı: 3.0.1
 title: NBA API
 description: "API for retrieving NBA teams, matches, players, and
   statistics"
 version: 1.0.0
paths:
 /teams:
 /match/{match id}/stats:
     summary: Retrieve match stats by match ID
     parameters:
       - name: match id
         in: path
         description: Match ID
         required: true
         schema:
          type: string
         example: "0022301187"
     responses:
       "200":
        description: Match stats
           application/json:
             schema:
              $ref: '#/components/schemas/MatchStats'
 /players/{player id}/stats:
 /standings:
 /feature vector:
 schemas:
  Standings: 📟
   TeamStandings:
   Team: 
   TeamInfo:
```







bet_api



bet_api

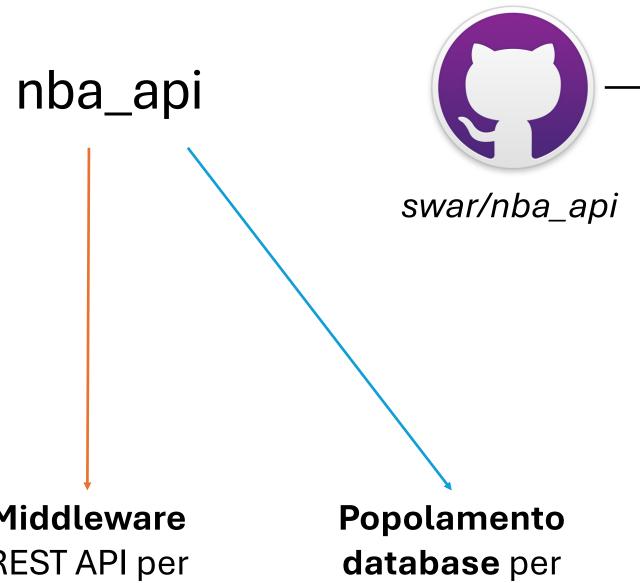
```
@Configuration
public class ClientConfiguration {
    @Value("${odds_api.url}")
    public String baseUrl;
    @Value("${odds_api.key}")
    public String apiKey;
    ♣ Leonardo Ruberto *
    @Bean
    public RestClient restClient() {
        return RestClient.builder()
                .requestFactory(new HttpComponentsClientHttpRequestFactory())
                .baseUrl(baseUrl)
                .requestInterceptor(new ApiKeyInterceptor(apiKey))
                .build();
```



- Inversion of Control
- Annotazioni

RestClient

Interceptor



Middleware REST API per applicazione web

Popolamento
database per
allenare modello di
Machine Learning

```
/match/{match_id}/stats:
 get:
   summary: Retrieve match stats by match ID
   operationId: get_match_stats
   parameters:
     - name: <u>match_id</u>
       in: path
       description: Match ID
       required: true
       style: simple
       explode: false
       schema:
         type: string
       example: "0022301187"
   responses:
     "200":
       description: Match stats
       content:
         application/json:
           schema:
             $ref: '#/components/schemas/MatchStats'
   x-openapi-router-controller: controllers.match_controller
```

Il Dataset

La creazione del Dataset risulta la vera prima challenge, poiché senza i dati consoni non è possibile ottenere un modello avente una accuracy soddisfacente

Sono stati analizzati vari articoli online tra cui "Which NBA Statistic Actually Translate to Wins" by Chinmay Vayda e parecchie repository di GitHub tra cui "Analysis of the Outcome of NBA Games" redatto dalla Bryant University

```
['home_team', 'away_team', 'game_id', 'season', 'date', 'pts_H', 'pts_A', 'winner', 'playoff',
'fg_pct_A', 'fg3_pct_A', 'ft_pct_A', 'reb_A', 'tot_ast_A', 'tot_tov_A', 'tot_stl_A',
'tot_blk_A', 'ts_pct_A', 'w_pct_A', 'w_pct_last_5_games_A', 'fg_pct_H', 'fg3_pct_H',
'ft_pct_H', 'reb_H', 'tot_ast_H', 'tot_tov_H', 'tot_stl_H', 'tot_blk_H',
'ts_pct_H', 'w_pct_H', 'w_pct_last_5_games_H', 'off_rating_A', 'def_rating_A',
'off_rating_H', 'def_rating_H', 'lineup_efficiency_A', 'lineup_efficiency_H',
'referee_name', 'referee_id']
```

Non tutte queste metriche sono state alla fine utilizzate, alcune è stato necessario ignorarle per avere modelli più performanti e conseguentemente più accurati

Approccio al problema

Obiettivo: Creazione di un modello capace di predire il risultato delle partite NBA

Risoluzione:

MODELLO DI CLASSIFICAZIONE

Classificazione binaria

Multi-Layer Perceptron

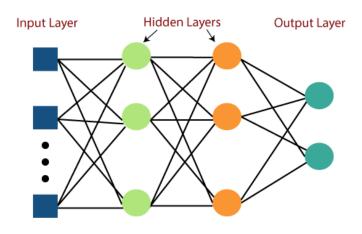
MODELLO DI CLASSIFICAZIONE

Previsione dell'esito esatto

Random Forest
Support Vector Machine
XGBOOST
Regressione Lineare
K-Nearest Neighbors
Bayesian

Modello di Classificazione

Il Multi-Layer Perceptron è una classe di reti neurali artificiali ampiamente utilizzata nei problemi di classificazione



```
model = Sequential([
    Dense(32, activation='relu'),
    Dropout(0.3),
    Dense(16, activation='relu'),
    Dense(1, activation='sigmoid')
])
```

Modello Sequenziale creato:

- 3 Layer Dense
- Layer Dropout per prevenire overfitting

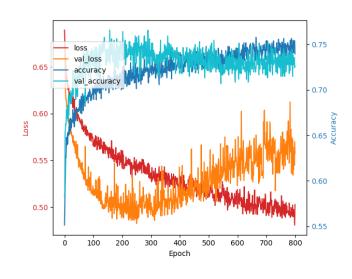
Risultati Modello di Classificazione

Results MLP:

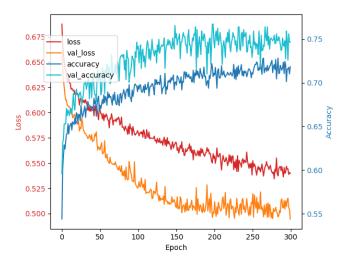
• Loss: 0.5629295706748962

Accuracy: 0.6992385983467102

Risultati con ricerca nello spazio degli iperparametri e con fine-tuning manuale per ottenere il risultato migliore:



Overfitting



Ottimizzato

Esempi di competitor presenti online: FiveThirtyEight (76%), Bryant Univeristy (65%), average (62-65%)

Modello di Regressione

Il problema di classificazione è stato convertito in un problema di regressione, ovvero previsione della differenza dei punti tra le squadre. La previsione ci fornisce anche il vincitore, infatti è calcolato come:

data['point_diff'] = data['pts_H'] - data['pts_A']

Dunque se il valore è positivo ha vinto la squadra di casa altrimenti ha vinto la squadra ospite

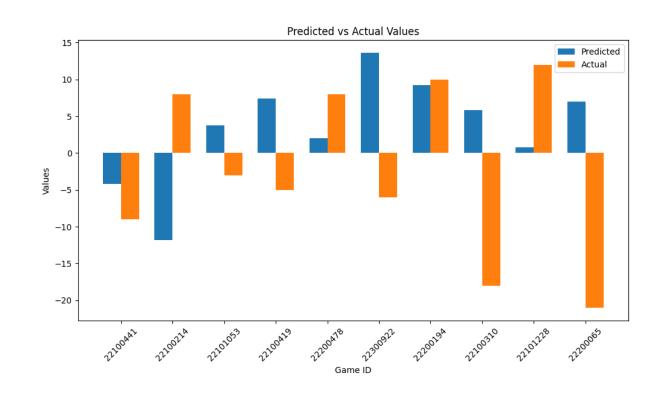
Metriche utilizzate per analisi e valutazione dei modelli creati

- Mean Absolute Error (MAE): Il MAE misura la deviazione media tra le predizioni del modello e i valori reali, ovvero indica l'errore medio tra la differenza di punti predetti dal modello e i risultati reali delle partite
- Mean Squared Error (MSE): Il MSE misura la media dei quadrati delle differenze tra predizioni e valori reali
- Root Mean Square Error (RMSE): Il RMSE è la radice quadrata del MSE ed è espresso nelle stesse unità del target
- **Sign Accuracy**: La Sign Accuracy rappresenta la precisione del modello adattando il modello di regressione ad uno di classificazione. Calcolato come: if (segnoConcorde): previsione vincitrice esatta

I modelli sono esposti in ordine crescente di accuracy e ognuno di loro è stato fine-tunato tramite una GridSearch per trovare la combinazione di iperparametri che li facesse performare meglio

K-Nearest Neighbors

Il KNN è un algoritmo di machine learning che classifica i dati nuovi basandosi sui dati più vicini nel dataset di addestramento.In pratica, trova i "K" punti più vicini al nuovo punto e calcola la media dei loro valori per fare la previsione, utilizzando così i vicini più simili per stimare il valore continuo del nuovo dato.

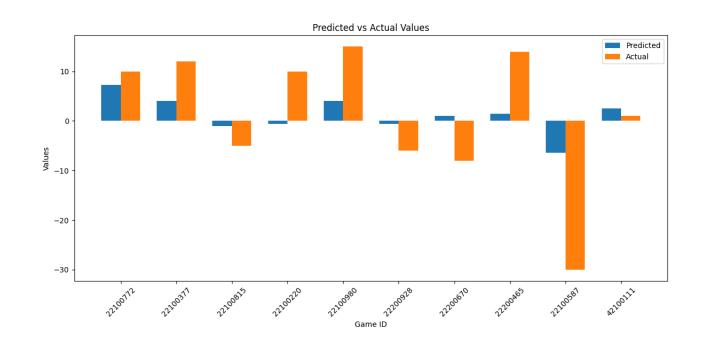


Results of KNN:

Mean Absolute Error (MAE): 11.620278833967047 Mean Squared Error (MSE): 215.66732572877058

Root Mean Squared Error (RMSE): 14.6856162869922

Support Vector Machine



Support Vector Machine per la Regressione (SVR) è un algoritmo che cerca di trovare una funzione che approssimi al meglio la relazione tra le variabili di input e output. Invece di cercare di passare esattamente attraverso tutti i punti dati, SVR permette a qualche punto di cadere all'interno di una "zona di tolleranza" intorno alla funzione stimata, cercando di minimizzare gli errori complessivi ma permettendo alcune deviazioni.

Results of SVM:

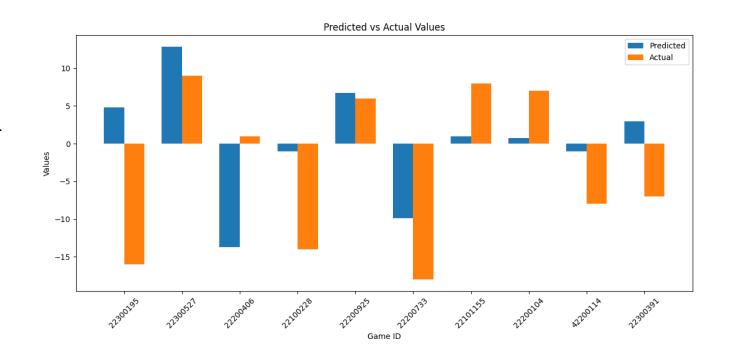
Mean Absolute Error (MAE): 10.360566704859455

Mean Squared Error (MSE): 173.73226432418772

Root Mean Squared Error (RMSE): 13.18075355676555

Random Forest

Random Forest è un algoritmo di machine learning che crea molti alberi di decisione indipendenti durante l'addestramento. Per fare una previsione, ciascun albero di decisione nel "bosco" (forest) fornisce una stima del valore continuo. L'algoritmo calcola la media di tutte queste stime ottenendo una previsione più accurata e robusta rispetto a quella che potrebbe dare un singolo albero. Questo approccio aiuta a ridurre il rischio di overfitting e migliora la generalizzazione del modello.



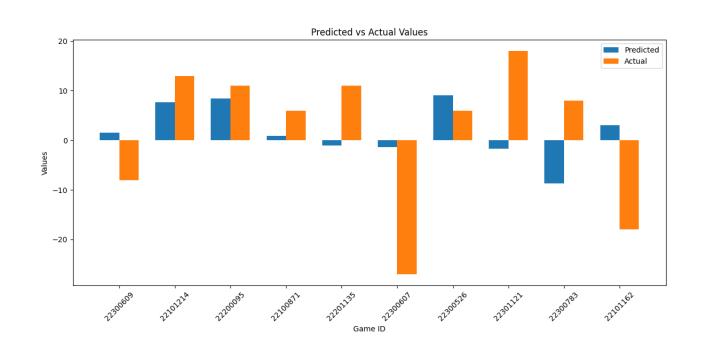
Results of RF:

Mean Absolute Error (MAE): 9.944223032615545

Mean Squared Error (MSE): 159.938928930258

Root Mean Squared Error (RMSE): 12.646696364278617

XGBoost



XGBoost è un algoritmo di machine learning basato sul boosting, una tecnica che crea un modello forte combinando molti modelli deboli. In XGBoost per la regressione, ogni nuovo modello viene costruito per correggere gli errori fatti dai modelli precedenti. Questo processo di aggiungere modelli uno dopo l'altro permette di migliorare continuamente la precisione della previsione. XGBoost utilizza anche tecniche di regolarizzazione per prevenire l'overfitting, rendendolo molto efficace e robusto per problemi di regressione.

Results of XGBoost:

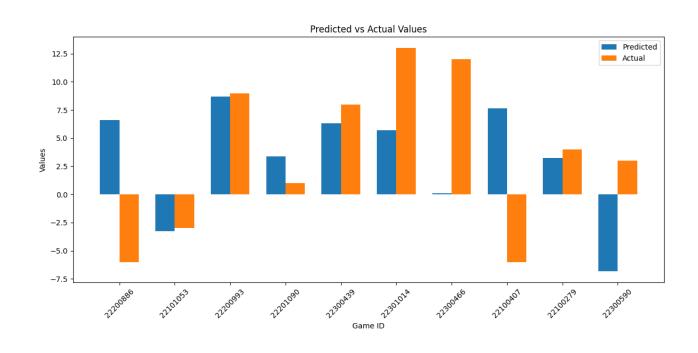
Mean Absolute Error (MAE): 9.6111730557072

Mean Squared Error (MSE): 149.89162676035912

Root Mean Squared Error (RMSE): 12.243023595515902

Bayesian

La regressione bayesiana è un metodo di machine learning che utilizza il teorema di Bayes per aggiornare le previsioni sulla base di nuove informazioni. Inizialmente, si parte con una stima preliminare (distribuzione a priori) della relazione tra variabili di input e output. Man mano che si ottengono nuovi dati, questa stima viene aggiornata (distribuzione a posteriori), incorporando le nuove evidenze per migliorare la previsione. Questo approccio permette di gestire l'incertezza nei dati e nei modelli, offrendo previsioni più affidabili e interpretabili.

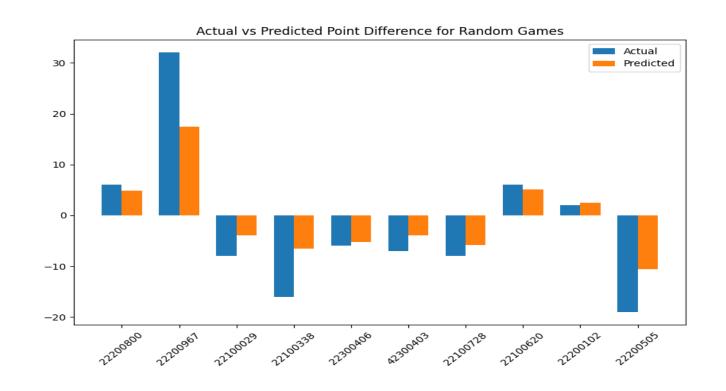


Results of Bayesian:

Mean Absolute Error (MAE): 9.684169695449112 Mean Squared Error (MSE): 151.8468062429373

Root Mean Squared Error (RMSE): 12.322613612498662

Linear Regression



La regressione lineare è un algoritmo di machine learning che modella la relazione tra una variabile dipendente (target) e una o più variabili indipendenti (feature) come una linea retta. L'obiettivo è trovare la linea che meglio approssima i dati minimizzando la somma degli errori quadrati tra i valori osservati e quelli predetti. Questo modello permette di fare previsioni continue, dove il valore predetto per un nuovo punto si basa su una combinazione lineare delle sue caratteristiche.

Results of Linear Regression:

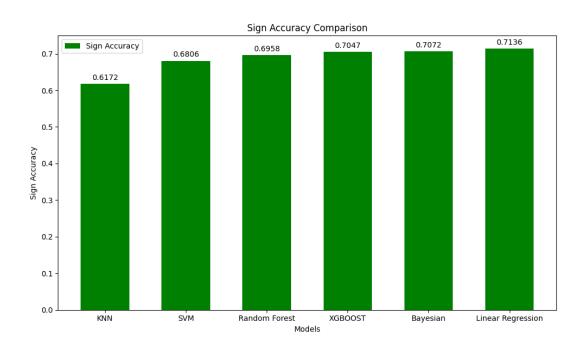
Mean Absolute Error (MAE): 9.673524553632301

Mean Squared Error (MSE): 151.58868540348476

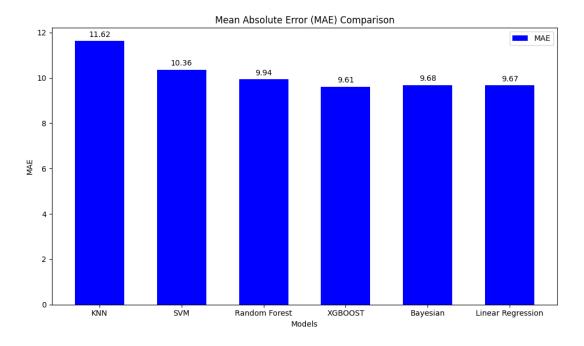
Root Mean Squared Error (RMSE): 12.312135696274824

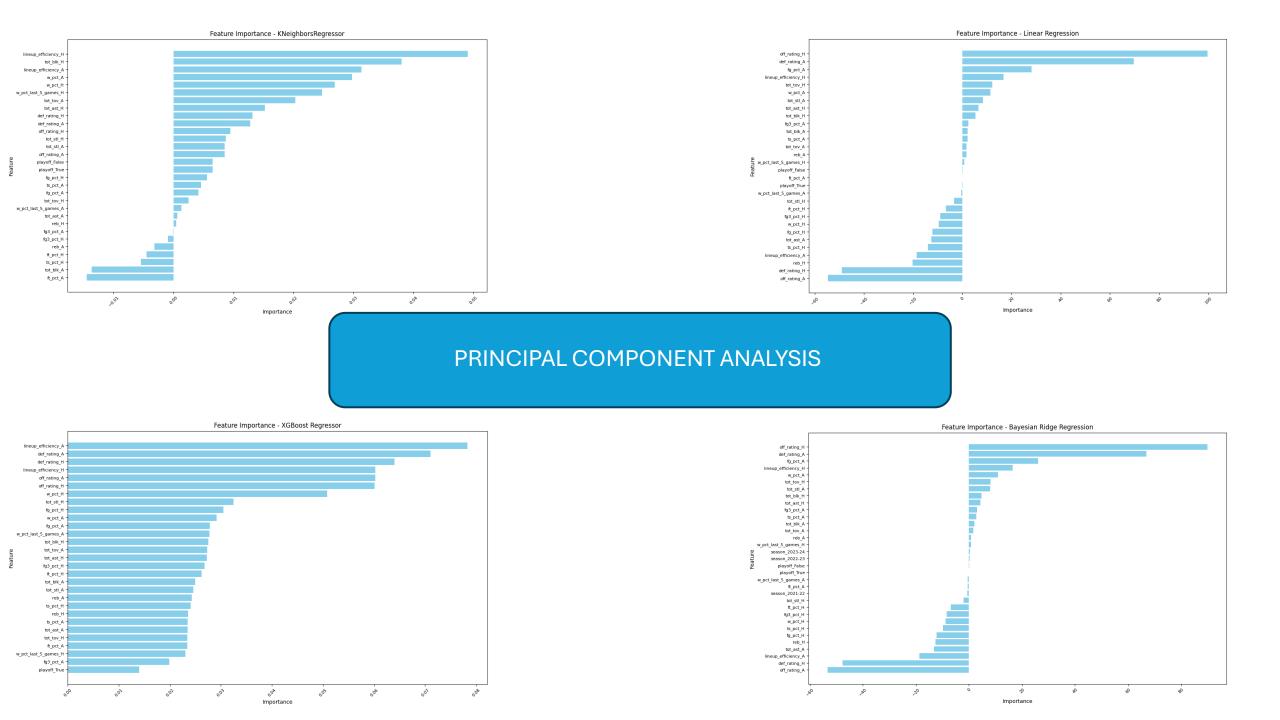
Risultati

Risultati del problema di Classificazione:

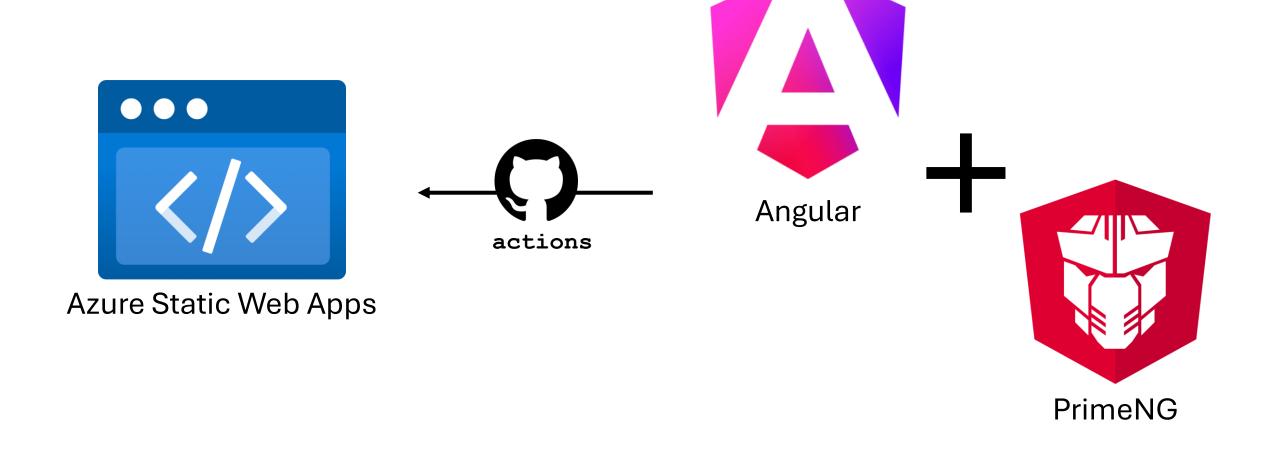


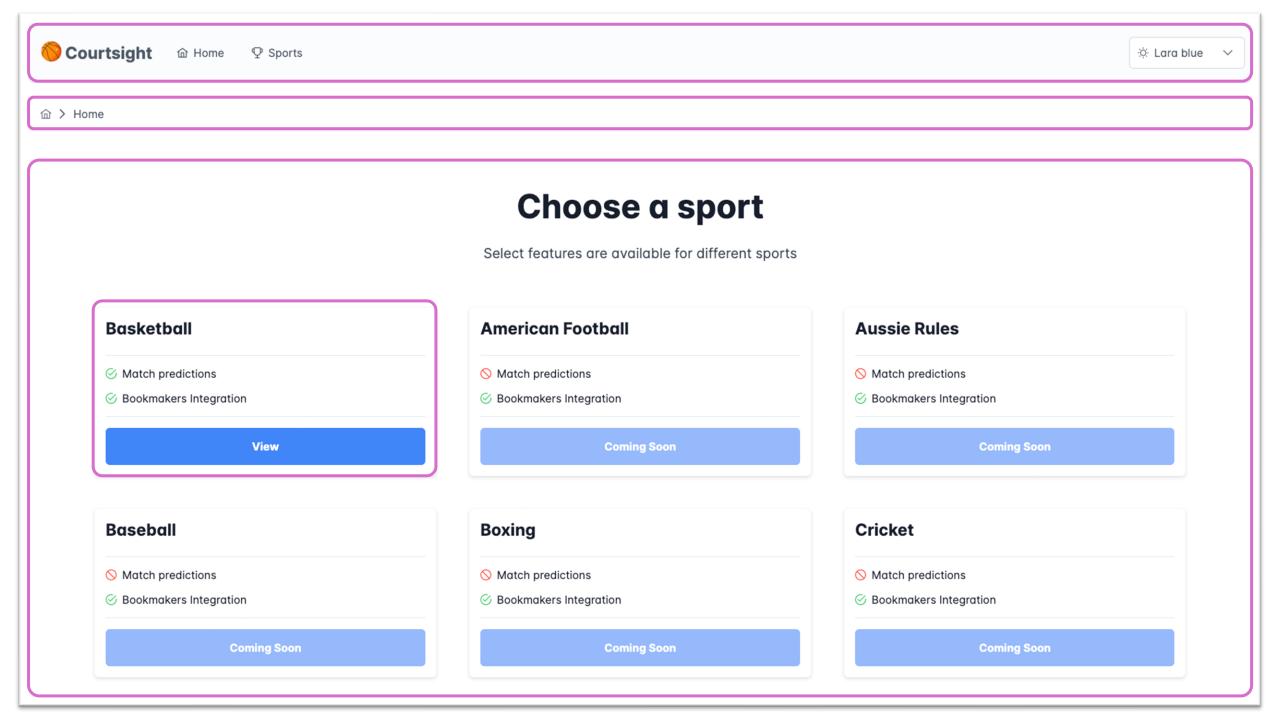
Risultati del problema di Classificazione:

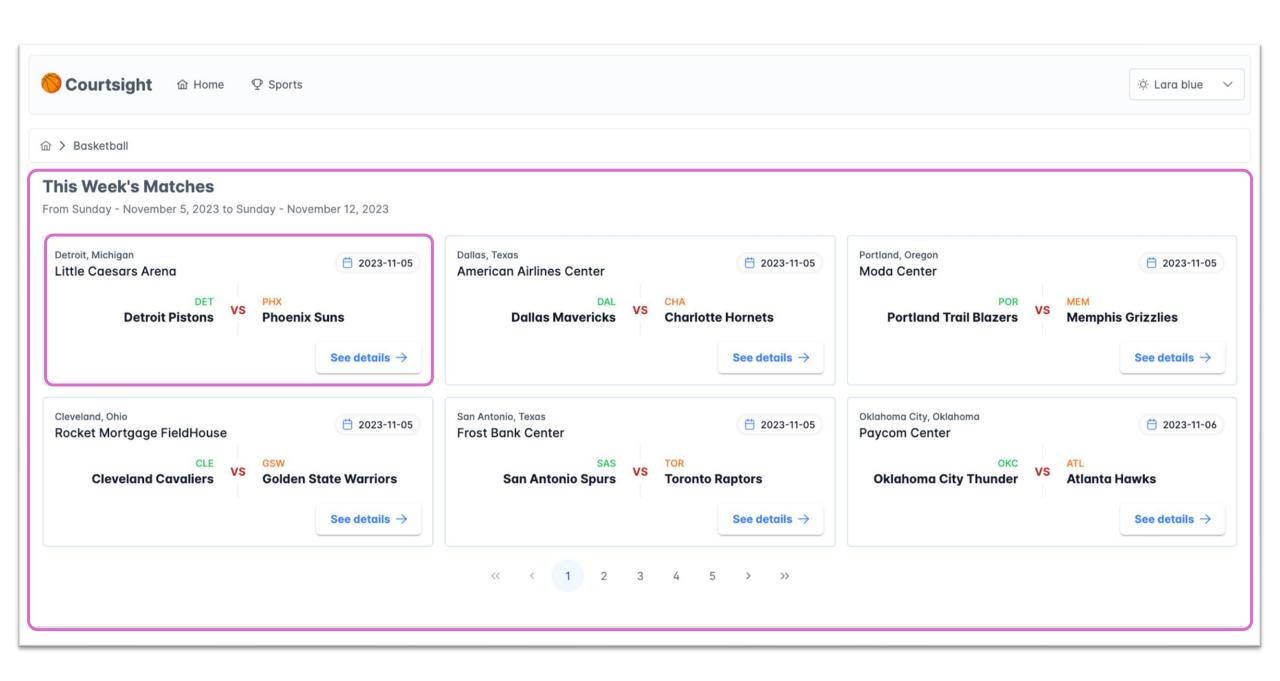




Web App







Standings

East Coast

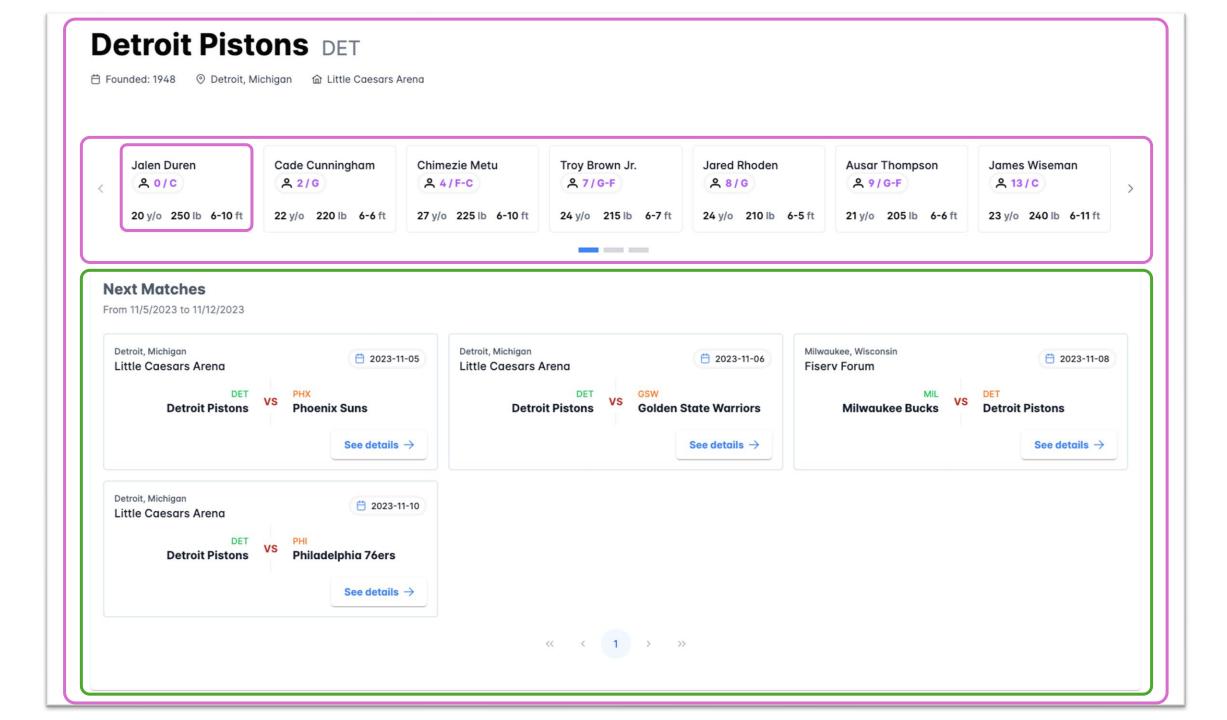
#	Team	Games	W	L	W %	Home record	Road record
1	Boston Celtics	5	5	0	1	2-0	3-0
2	Philadelphia 76ers	5	4	1	0.8	3-0	1-1
3	Atlanta Hawks	6	4	2	0.667	2-1	2-1
4	Orlando Magic	6	4	2	0.667	2-0	2-2
5	Milwaukee Bucks	5	3	2	0.6	3-1	0-1
6	Brooklyn Nets	6	3	3	0.5	0-2	3-1
7	Indiana Pacers	6	3	3	0.5	2-2	1-1
8	Cleveland Cavaliers	7	3	4	0.429	1-3	2-1
9	Toronto Raptors	7	3	4	0.429	2-2	1-2
10	Charlotte Hornets	6	2	4	0.333	1-2	1-2
11	Miami Heat	6	2	4	0.333	2-1	0-3
12	New York Knicks	6	2	4	0.333	0-2	2-2
13	Chicago Bulls	7	2	5	0.286	1-2	1-3
14	Detroit Pistons	7	2	5	0.286	1-2	1-3
15	Washington Wizards	5	1	4	0.2	1-1	0-3

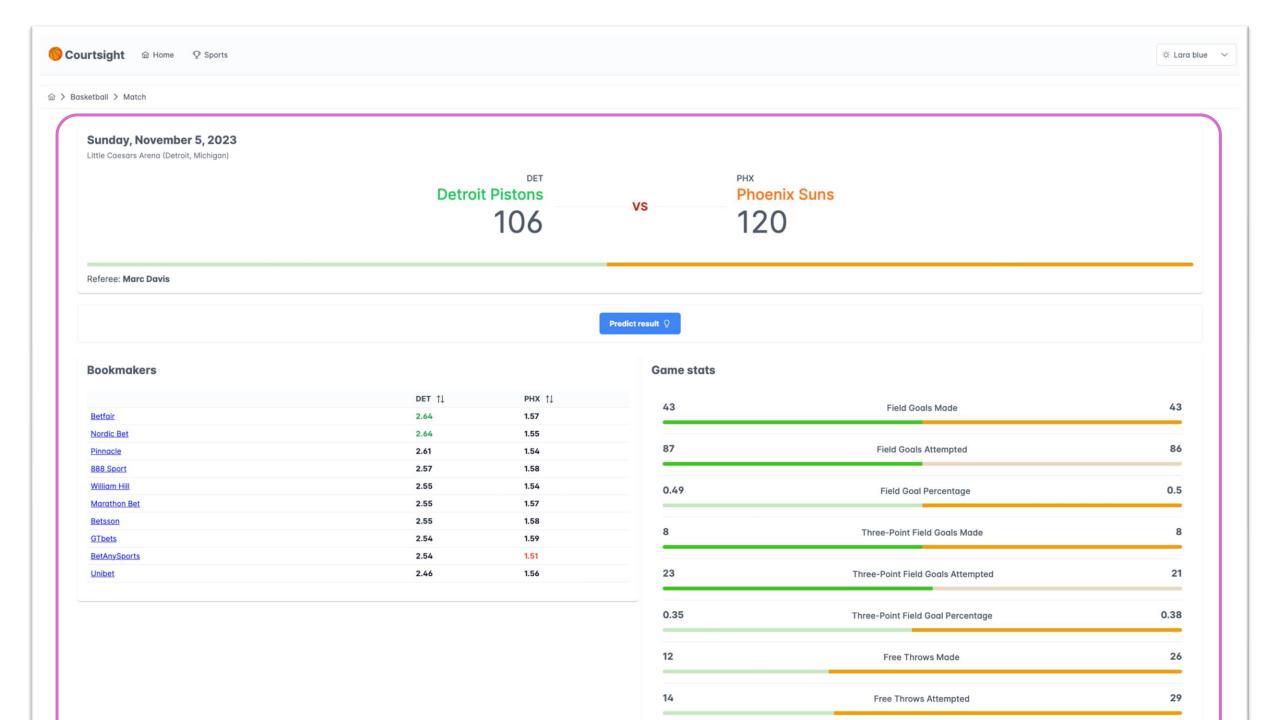
West Coast

#	Team	Games	w	L	W %	Home record	Road record
1	Denver Nuggets	7	6	1	0.857	4-0	2-1
2	Dallas Mavericks	6	5	1	0.833	3-0	2-1
3	Golden State Warriors	7	5	2	0.714	1-1	4-1
4	New Orleans Pelicans	6	4	2	0.667	2-2	2-0
5	Los Angeles Clippers	5	3	2	0.6	3-0	0-2
6	Minnesota Timberwolves	5	3	2	0.6	3-0	0-2
7	Los Angeles Lakers	6	3	3	0.5	3-0	0-3
8	Oklahoma City Thunder	6	3	3	0.5	1-3	2-0
9	San Antonio Spurs	6	3	3	0.5	1-2	2-1
10	Phoenix Suns	7	3	4	0.429	1-2	2-2
11	Portland Trail Blazers	7	3	4	0.429	1-2	2-2
12	Houston Rockets	5	2	3	0.4	2-1	0-2
13	Sacramento Kings	5	2	3	0.4	1-1	1-2
14	Utah Jazz	7	2	5	0.286	2-2	0-3
15	Memphis Grizzlies	7	1	6	0.143	0-3	1-3

Playoff Playin

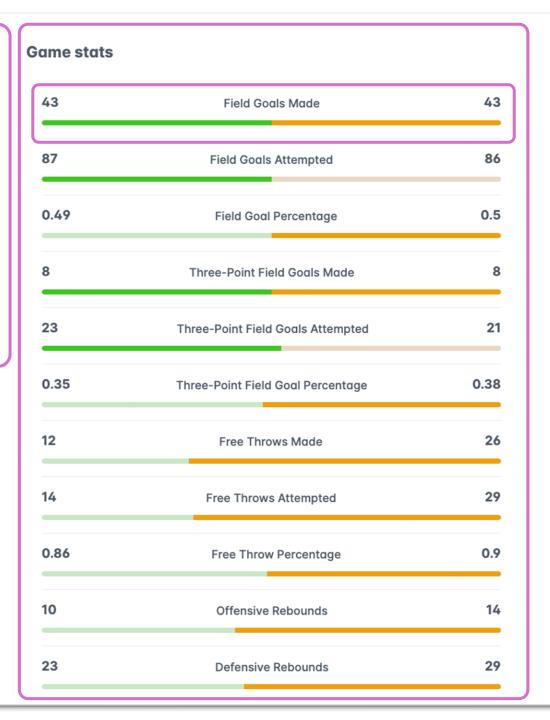
Last update: 11/5/23, 1:00

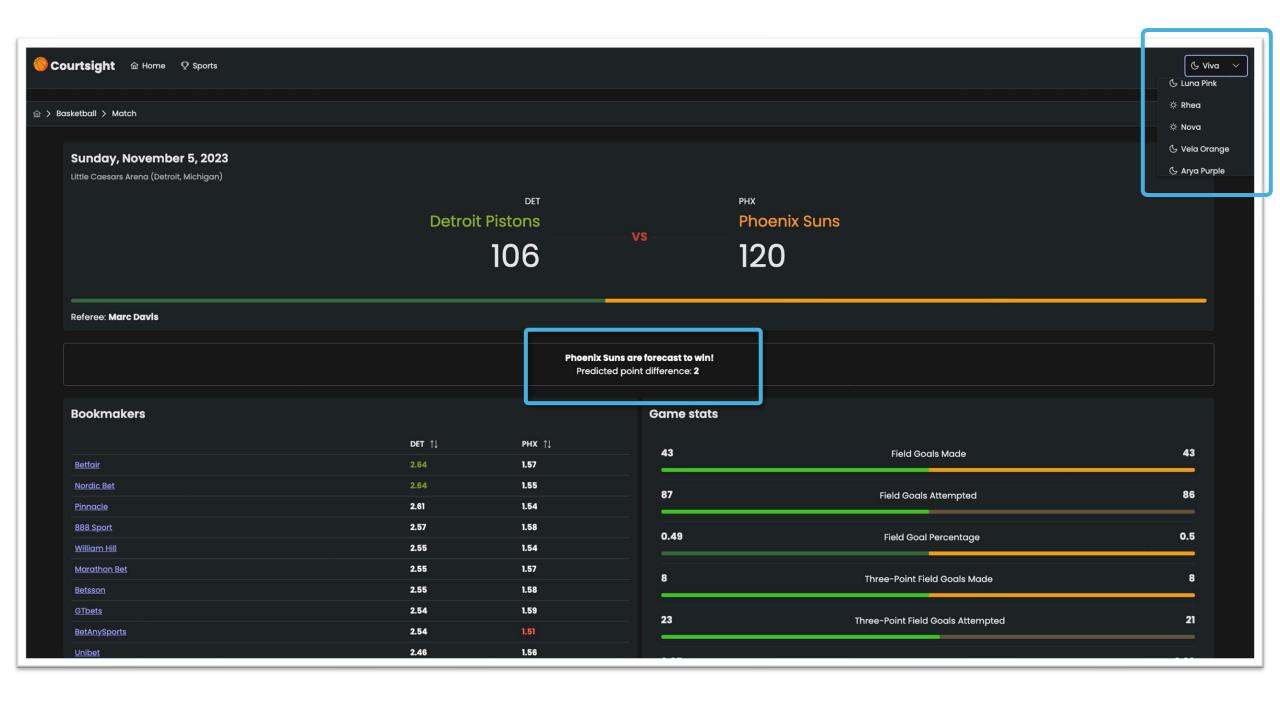




Bookmakers

Nordic Bet 2.64 1.55 BetAnySports 2.62 1.55 Unibet 2.61 1.58 Pinnacle 2.60 1.56 Marathon Bet 2.56 1.52 Betfair 2.54 1.50 GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60 888 Sport 2.48 1.56		DET ↑↓	PHX ↑↓
Unibet 2.61 1.58 Pinnacle 2.60 1.56 Marathon Bet 2.56 1.52 Betfair 2.54 1.50 GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60	Nordic Bet	2.64	1.55
Pinnacle 2.60 1.56 Marathon Bet 2.56 1.52 Betfair 2.54 1.50 GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60	<u>BetAnySports</u>	2.62	1.55
Marathon Bet 2.56 1.52 Betfair 2.54 1.50 GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60	<u>Unibet</u>	2.61	1.58
Betfair 2.54 1.50 GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60	<u>Pinnacle</u>	2.60	1.56
GTbets 2.52 1.51 Betsson 2.50 1.53 William Hill 2.49 1.60	Marathon Bet	2.56	1.52
Betsson 2.50 1.53 William Hill 2.49 1.60	Betfair	2.54	1.50
William Hill 2.49 1.60	<u>GTbets</u>	2.52	1.51
	Betsson	2.50	1.53
888 Sport 2.48 1.56	William Hill	2.49	1.60
	888 Sport	2.48	1.56





East	Coast									West	Coast									
#	Team	Games	w	L	W %	Home record	Road reco	ord		#	Tean	n			Games	w	L	W %	Home record	Road record
1	Boston Celtics	72	57	15	0.792	32-3	25-12			1	Denv	er Nugg	ets		73	51	22	0.699	29-7	22-15
2	Milwaukee Bucks	72	46	26	0.639	29-8	17-18			2	Minn	esota Ti	mberw	olves	72	50	22	0.694	26-9	24-13
3	New York Knicks	72	44	28	0.611	24-13	20-15			3	Okla	homa Ci	ty Thur	nder	72	50	22	0.694	28-8	22-14
4	Cleveland Cavaliers	73	44	29	0.603	23-14	21-15			4	Los	Angeles	Clipper	s	72	45	27	0.625	22-13	23-14
5	Orlando Magic	72	42	30	0.583	25-11	17-19			5	New	Orleans	Pelicar	ns	72	44	28	0.611	20-14	24-14
6	Indiana Pacers	74	41	33	0.554	21-15	20-18			6	Dallo	ıs Maver	icks		72	43	29	0.597	22-15	21-14
7	Miami Heat	72	39	33	0.542	18-17	21-16			7	Phoe	nix Suns	5		73	43	30	0.589	23-14	20-16
8	Philadelphia 76ers	73	39	34	0.534	21-16	18-18			8	Sacr	amento	Kings		72	42	30	0.583	21-14	21-16
9	Chicago Bulls	73	35	38	0.479	19-19	16-19			9	Los A	Angeles	Lakers		73	41	32	0.562	27-12	14-20
10	Atlanta Hawks	72	33	39	0.458	19-17	14-22	<		Ма	rch 2	2024		>	72	38	34	0.528	18-19	20-15
11	Brooklyn Nets	73	28	45	0.384	16-19	12-26			IVIG		-02-4			72	37	35	0.514	26-11	11-24
12	Toronto Raptors	73	23	50	0.315	13-24	10-26	Su	Мо	Tu	We	Th	Fr	Sa	73	29	44	0.397	20-17	9-27
13	Charlotte Hornets	72	18	54	0.25	10-24	8-30	25	26	27	28	29	1	2	73	24	49	0.329	8-28	16-21
14	Washington Wizards	73	14	59	0.192	6-29	8-30			-		_			73	19	54	0.26	11-27	8-27
15	Detroit Pistons	73	12	61	0.164	7-31	5-30	3	4	5	6	7	8	9	73	17	56	0.233	9-27	8-29
								10	11	12	13	14	15	16						
Play	off Playin							17	18	19	20	21	22	23						
upda	ite: 3/28/24, 0:00							24	25	26	27	28	29	30						
								31	1	2	3	4	5	6						

DEPLOYMENT



COMPONENTI

Componenti	Servizi
Middleware Bet Api	Azure Container App
Middleware NBA Api	Azure Container App
Frontend Applicazione Web	Azure Static Web App
Modello Machine Learning	Azure Machine Learning







SERVIZI AGGIUNTIVI

GESTIONE IMMAGINI OCI

GESTIONE DNS

GESTIONE LOG

DOCUMENT DATABASE

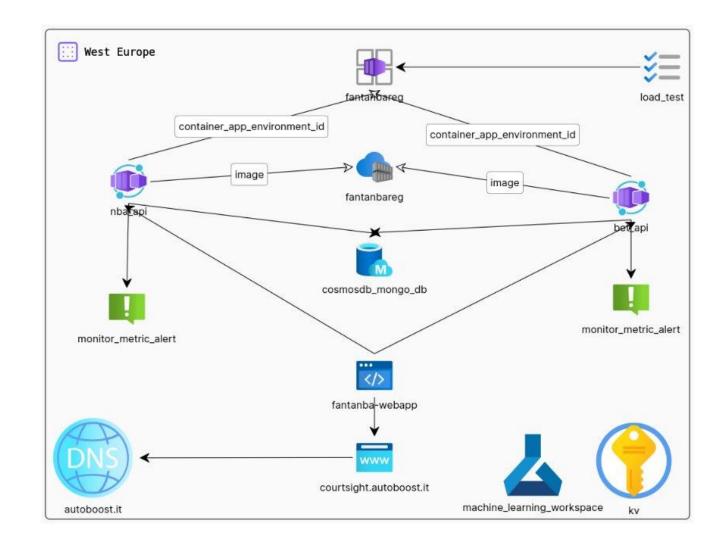




ARCHITETTURA FINALE

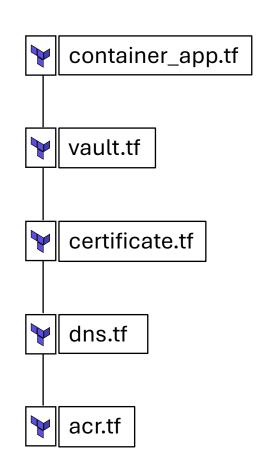
L'architettura include:

- sistemi di raccolta e analisi dei log
- sistema di alert su metriche
- DNS zone per gestione dominio



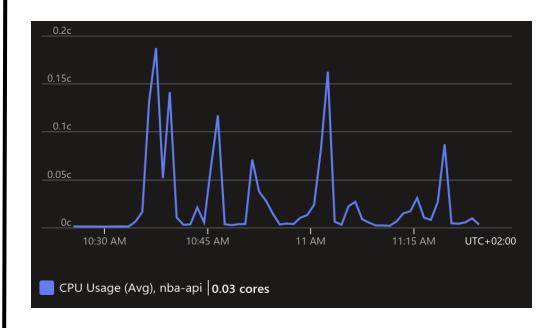
PROVISIONING

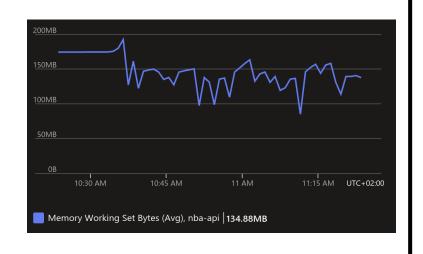


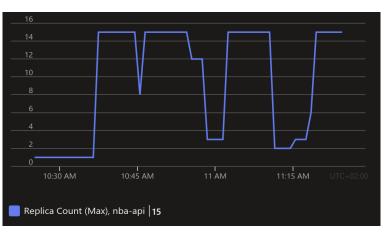


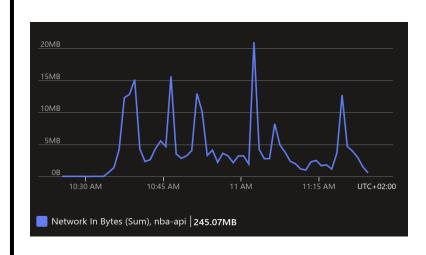
terraform apply
terraform destroy

LOAD TESTING NBA API

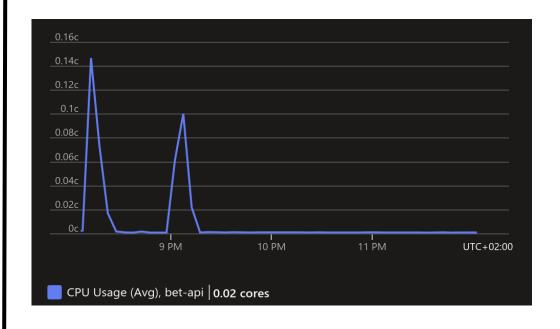


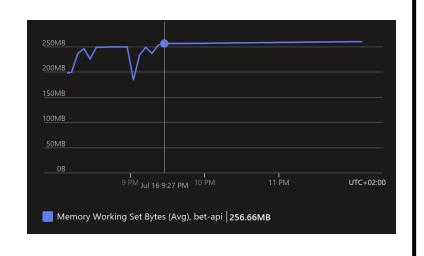


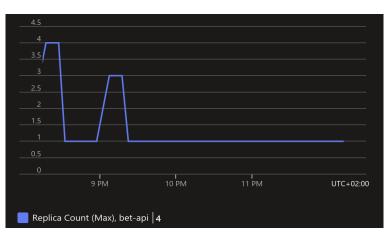


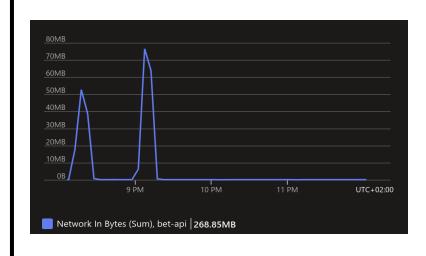


LOAD TESTING BET API









ANALISI DEI COSTI

A cause di problematiche con l'accesso alla dashboard Cost Analysis i costi riportati sono una stima a partire dai prezzi dichiarati e i carichi misurati empiricamente.

Services	€/Mese
Azure cosmos DB	~6
Load Test	~ 10 (50h test)
Azure ML Workspace	~ 90
Container Apps x2	~ 600 (3x load test)
Registry x2	~ 10
DNS Zone	~ 0.80 (1 mld richieste)
Azure key vault	~ 5
Log Analytics x2	~ 28
Various Alerts and Metrics	~0.00
Static Web App	~ 80 (3x load test)
Totale	830