# Sistem de recomandare universitate

Călugărițoiu Ion-Victor Rizea Adelina-Maria Albuică Sergiu-Mihail

IBD 2025

# Funcționalități principale

## Jupyter Notebooks

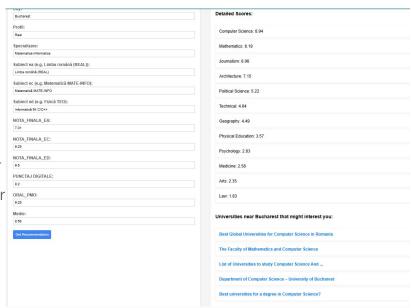
- 1. Dataset cleaning
- 2. Analiza rezultatelor de la Bacalaureat folosind algoritmi de ML Random Forest Classifier, K-Means Clustering
- 3. Vizualizarea dataset-urilor de la admiterile de la universități histograme, boxplots, scatterplots, bar/pie charts
- 4. Recomandarea unei facultăți relevante în funcție de profilul unui candidat

# Web App - Microsoft Azure

## Funcționalități:

- Introducere profil unui candidat la Bacalaureat (rezultate, profil, specializare, etc)
- Recomandare profil facultate în funcție de datele introduse
- Comparație între profilul introdus și cele din dataset-ul rezultatelor de la Bacalaureat
- Vizualizare date relevante din rezultatele dataset-urilor admiterilor

Aplicația web și Notebook-ul Jupyter rulează pe Azure.



## Seturi de date

Rezultate examen Bacalaureat

 ⊞ bac2019.csv
 5 hours ago

 ⊞ probe\_profil.csv
 5 hours ago

• Rezultate admiteri

| 🗽 ARHITECTURA_2024.pdf   | 5 hours ago |
|--------------------------|-------------|
| 🗽 DREPT_2024.pdf         | 5 hours ago |
| 🗽 FIZICA_2023.pdf        | 5 hours ago |
| 🗞 GEOGRAFIE_2024.pdf     | 5 hours ago |
| 🔖 PSIHOLOGIE_2024.pdf    | 5 hours ago |
| 🗞 SNSPA_2024.pdf         | 5 hours ago |
| SOCIOLOGIE_2024.pdf      | 5 hours ago |
| 🗽 UMFCD_2019.pdf         | 5 hours ago |
| unibuc_cti_info_2024.pdf | 5 hours ago |
| 🗽 UNIBUC_MATE_2024.pdf   | 5 hours ago |
| VPB_ACS_2014.pdf         | 5 hours ago |

# **Data integration & cleaning**

- Ștergerea coloanelor goale
- Eliminarea coloanelor nefolositoare
- Îmbinarea datelor
- Standardizarea datelor într-un format convenabil

|   | Nr.        | Dosar | Medie BAC | Prezenta | Nota Admitere | Medie Generala | Status      | Observatii | Loc repartizat |
|---|------------|-------|-----------|----------|---------------|----------------|-------------|------------|----------------|
| ( | 0 1        | 0084  | 9.63      | Prezent  | Olimpic       | Olimpic        | Admis Buget | INSCRIS    | B-1            |
|   | 1 2        | 0859  | 9.31      | Prezent  | 9.04          | 9.12           | Admis Buget | INSCRIS    | B-2            |
| 1 | <b>2</b> 3 | 0347  | 9.73      | Prezent  | 8.64          | 8.97           | Admis Buget | INSCRIS    | B-3            |
|   | 3 4        | 0168  | 9.51      | Prezent  | 8.46          | 8.78           | Admis Buget | INSCRIS    | B-4            |
|   | 4 5        | 0098  | 9.86      | Prezent  | 8.26          | 8.74           | Admis Buget | INSCRIS    | B-5            |

| ■ ARHITECTURA_clean_2024.csv | 5 hours ago |
|------------------------------|-------------|
| ☐ DREPT_clean_2024.csv       | 5 hours ago |
| ⊞ GEOGRAFIE_2024.csv         | 5 hours ago |
| ☐ JURNALISM_2024.csv         | 5 hours ago |
| ☐ PSIHOLOGIE_2024.csv        | 5 hours ago |
| ⊞ SNSPA_2024.csv             | 5 hours ago |
| ⊞ UMFCD_2019.csv             | 5 hours ago |
| ■ UNIBUC_CTI_INFO_2024.csv   | 5 hours ago |
| ■ UNIBUC_MATE_2024.csv       | 5 hours ago |
| Ⅲ UPB_ACS_2014.csv           | 5 hours ago |

|  |   | Nr. | Medie BAC | Nota Română |
|--|---|-----|-----------|-------------|
|  | 0 | 1   | 9.48      | 9.85        |
|  | 1 | 2   | 9.26      | 8.55        |
|  | 2 | 3   | 9.06      | 8.35        |
|  | 3 | 4   | 9.60      | 9.40        |
|  | 4 | 5   | 9.36      | 9.65        |

# **Bac Dataset - Algoritmi ML**

**Random Forest Classifier** pentru a prezice dacă un elev promovează sau nu Bacalaureatul, folosind dataset-ul rezultatelor de la Bacalaureat:

- Modelul este antrenat pe dataset
- Coloanele care indică notele probelor, status-ul PROMOVAT/NEPROMOVAT și media finală sunt ignorate pentru a evita influențarea modelului

#### Output:

- Se prezice statusul promovării și se afișează raportul de clasificare.
- Se afișează matricea de confuzie și un grafic cu feature importance.

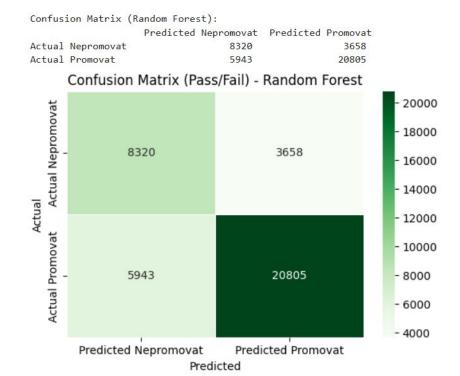
### **Random Forest Classifier** - rezultate

| Classificatio | n Report (Ra | ndom Fore | st):     |         |
|---------------|--------------|-----------|----------|---------|
|               | precision    | recall    | f1-score | support |
| Nepromovat    | 0.5833       | 0.6946    | 0.6341   | 11978   |
| Promovat      | 0.8505       | 0.7778    | 0.8125   | 26748   |
| accuracy      |              |           | 0.7521   | 38726   |

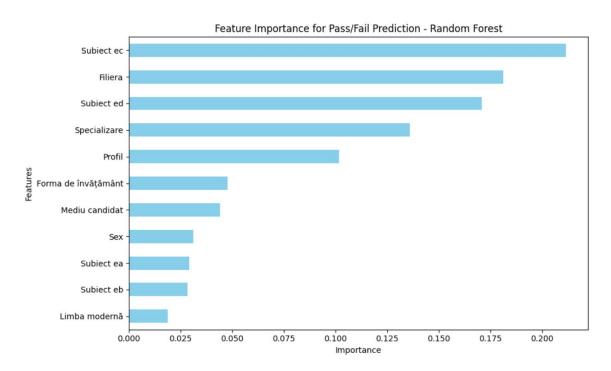
**Nepromovat**: **0.5833** → Doar 58.33% dintre elevii prezis ca *Nepromovat* au fost corect clasificați.

**Promovat**:  $0.8505 \rightarrow 85.05\%$  dintre elevii prezis ca *Promovat* au fost corect clasificați.

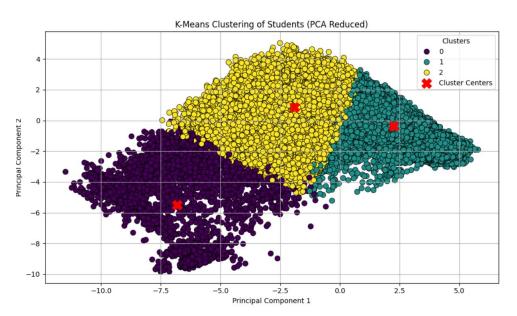
**Observație:** Modelul este mult mai precis în identificarea elevilor care promovează decât a celor care nu promovează.



## Random Forest Classifier - Importanța coloanelor



## K-Means Clustering - grupare pe baza similarităților elevilor



#### Cluster Summary:

Cluster 0: 5169 students (3.80% of total) Cluster 1: 68526 students (50.35% of total) Cluster 2: 62396 students (45.85% of total)

Top Specializations in Each Cluster:

#### Cluster 0:

- Tehnician mecanic pentru întreținere și reparații: 390 students
- Științe ale Naturii: 362 students
- Filologie: 288 students

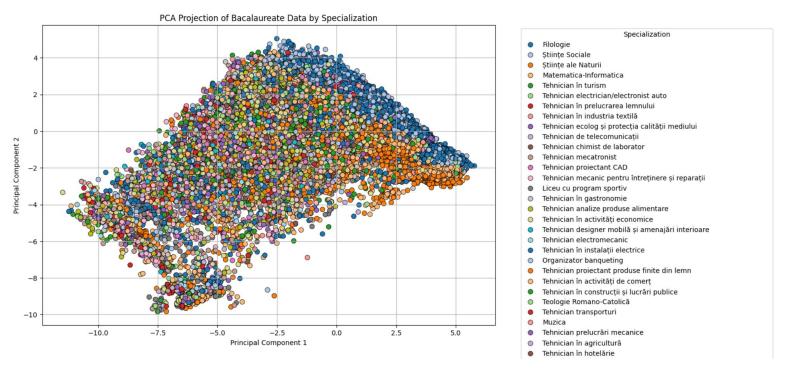
#### Cluster 1:

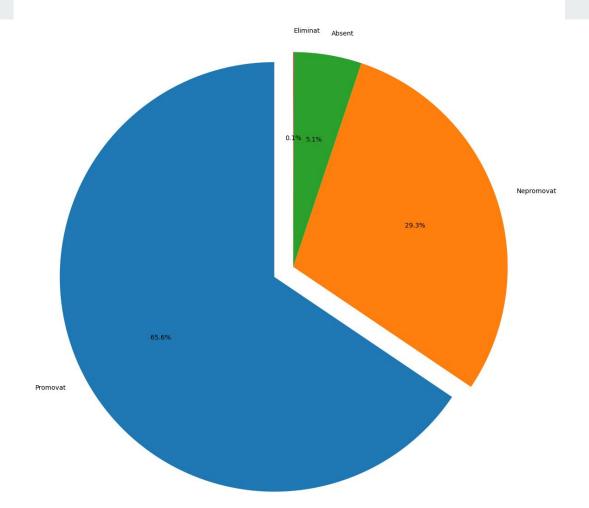
- Matematica-Informatica: 20245 students
- Filologie: 15064 students
- Științe ale Naturii: 12344 students

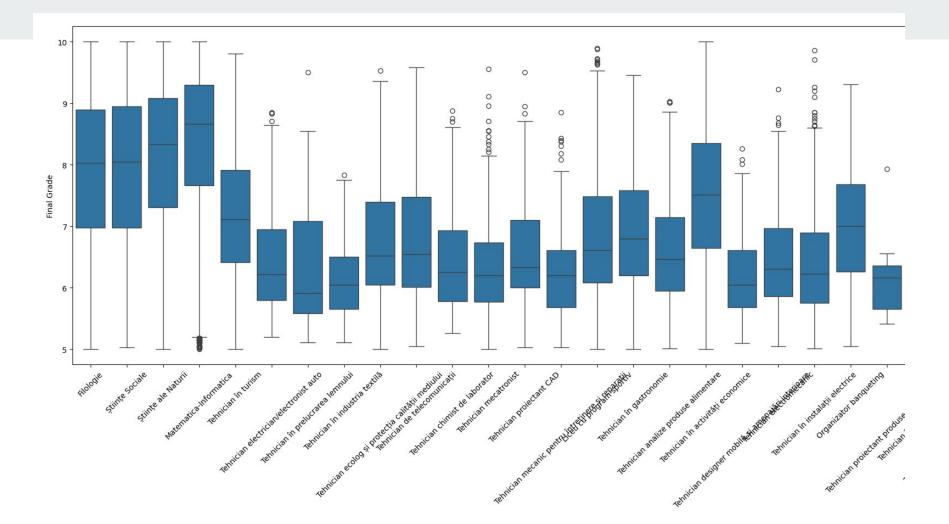
#### Cluster 2:

- Științe ale Naturii: 7200 students
- Tehnician în activități economice: 6328 students
- Filologie: 6149 students

## Principal Component Analysis (PCA) - grupare/specializare







Correlation Heatmap of Numeric Features Cod unic candidat - 1.00 -0.07 0.01 0.11 0.11 0.02 0.08 -0.01 -0.02 -0.12 0.13 0.01 -0.01 0.05 -0.09 0.03 0.01 0.07 0.07 0.18 0.17 0.19 0.20 0.19 0.11 0.22 0.17 0.16 0.02 1.00 -0.06 -0.05 0.02 0.04 -0.15 -0.01 -0.09 0.05 -0.06 -0.03 0.00 -0.05 -0.02 -0.08 -0.01 -0.06 -0.08 0.03 0.00 -0.03 0.00 0.00 0.05 -0.11 -0.15 -0.19 -0.04 -0.10 -0.14 -0.11 -0.15 Profil - 0.01 -0.06 1.00 -0.33 -0.08 -0.05 0.05 0.06 -0.16 0.32 -0.02 0.01 -0.04 0.07 -0.03 -0.01 -0.01 -0.03 -0.05 -0.06 -0.04 -0.03 -0.03 -0.09 -0.02 -0.05 -0.06 -0.06 -0.04 -0.03 -0.05 -0.06 -0.05 -0.06 -0.04 -0.03 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.06 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.05 -0.06 -0.05 -0.0 Filiera - 0.11 -0.05 -0.33 1.00 0.04 0.16 0.29 0.02 0.00 -0.79 0.24 0.06 -0.06 0.14 -0.12 0.16 0.07 0.30 0.23 0.21 0.21 0.24 0.20 0.20 0.18 0.30 0.15 0.27 0.04 0.38 0.32 0.45 0.09 Forma de învătământ - 0.11 0.02 -0.08 0.04 1.00 -0.01 -0.05 -0.00 -0.00 -0.00 -0.00 0.02 0.02 -0.02 0.07 -0.04 0.16 0.03 0.15 0.15 0.06 0.06 0.07 0.08 0.07 0.18 0.05 0.18 0.05 0.18 0.01 0.16 0.16 0.18 0.08 Mediu candidat - 0.02 0.04 -0.05 0.16 -0.01 1.00 0.05 -0.04 -0.02 -0.16 0.08 0.03 -0.01 0.17 -0.11 0.11 0.01 0.14 0.10 0.22 0.22 0.24 0.22 0.24 0.22 0.25 0.15 0.15 0.11 0.17 0.01 0.29 -0.13 Subject ea - 0.08 -0.15 0.70 0.29 -0.05 0.05 1.00 0.08 0.07 0.64 0.54 0.03 0.00 0.04 0.01 0.09 -0.00 0.18 0.13 0.08 0.07 0.12 0.10 0.09 0.01 0.18 0.05 0.12 -0.00 0.21 0.21 0.26 0.23 Subject eb - -0.01 -0.01 0.05 0.02 -0.00 -0.04 0.08 1.00 0.08 -0.05 0.03 -0.00 -0.04 0.08 -0.05 0.03 -0.04 -0.05 0.03 -0.04 -0.05 0.03 -0.04 0.02 -0.01 -0.06 -0.06 -0.06 -0.06 -0.06 -0.06 -0.03 -0.03 -0.03 -0.07 -0.07 -0.03 -0.03 -0.03 -0.03 -0.03 -0.05 0.07 -0.07 Subject ec - 0.12 0.05 -0.16 -0.79 -0.00 -0.16 -0.79 -0.00 -0.16 -0.64 -0.05 -0.03 1.00 -0.46 -0.06 0.02 -0.16 0.10 -0.16 -0.03 -0.31 -0.23 -0.23 -0.23 -0.23 -0.23 -0.23 -0.22 -0.17 -0.31 -0.14 -0.27 -0.02 -0.39 -0.33 -0.48 -0.14 Subject ed - 0.13 -0.06 0.32 0.24 0.02 0.08 0.54 0.03 0.03 -0.46 1.00 0.03 0.01 0.06 -0.00 0.10 -0.01 0.20 0.18 0.08 0.07 0.12 0.10 0.10 0.03 0.20 0.04 0.13 -0.01 0.23 0.26 0.27 0.22 STATUS A - 0.01 -0.03 -0.02 0.06 0.02 0.03 0.03 -0.00 0.01 -0.06 0.03 1.00 0.04 0.12 -0.03 0.06 0.03 0.10 0.09 0.04 0.05 0.07 0.04 0.04 0.06 0.11 0.13 0.12 -0.03 0.14 0.12 0.14 0.05 STATUS B - -0.01 0.00 0.01 -0.06 -0.02 -0.01 0.00 -0.04 -0.03 0.02 0.01 -0.04 1.00 -0.02 0.01 0.08 -0.92 -0.02 0.00 0.01 -0.02 0.02 0.02 0.03 0.02 0.02 0.03 0.02 0.03 0.09 -0.01 -0.01 0.01 0.01 STATUS C - 0.05 -0.05 -0.05 -0.04 0.14 0.07 0.17 0.04 -0.05 0.05 -0.16 0.06 0.12 -0.02 1.00 -0.13 0.28 0.05 0.28 0.27 0.40 0.42 0.37 0.35 0.35 0.18 0.32 0.30 0.39 0.08 0.38 0.34 0.51 -0.12 STATUS D - -0.09 -0.02 0.07 -0.12 -0.04 -0.11 0.01 0.03 0.02 0.10 -0.00 -0.03 0.01 -0.13 1.00 -0.04 -0.02 -0.10 -0.05 -0.33 -0.33 -0.33 -0.34 -0.34 -0.50 -0.16 -0.25 -0.21 -0.05 -0.25 -0.18 -0.37 0.35 STATUS EA - 0.03 -0.08 -0.03 0.16 0.16 0.11 0.09 -0.04 -0.00 -0.16 0.10 0.06 0.08 0.28 -0.04 1.00 -0.04 0.65 0.68 0.10 0.09 0.16 0.16 0.16 0.22 0.68 0.00 0.78 0.14 STATUS EB - 0.01 -0.01 -0.01 0.07 0.03 0.01 -0.00 0.02 0.03 -0.03 -0.03 -0.01 0.03 -0.92 0.05 -0.02 -0.04 1.00 0.06 0.06 0.01 0.01 0.04 0.01 0.00 0.04 0.02 -0.02 -0.09 0.15 0.04 0.04 0.03 0.00 STATUS EC - 0.07 -0.06 -0.01 0.30 0.15 0.14 0.18 -0.01 0.00 -0.31 0.20 0.10 -0.02 0.28 -0.10 0.65 0.06 1.00 0.69 0.17 0.16 0.23 0.21 0.21 0.28 0.82 0.00 STATUS ED - 0.07 -0.08 -0.03 0.23 0.15 0.10 0.13 -0.01 0.01 -0.23 0.18 0.09 -0.02 0.27 -0.05 0.68 0.06 0.69 1.00 0.12 0.10 0.16 0.16 0.15 0.23 0.75 0.00 TTA - 0.18 0.03 -0.05 0.21 0.06 0.22 0.08 -0.06 -0.06 -0.03 0.08 0.04 0.00 0.40 -0.33 0.10 0.01 0.17 0.12 1.00 0.73 0.70 0.66 0.66 0.41 0.25 0.36 0.35 0.06 0.39 0.30 SCRIS ITC - 0.17 0.00 -0.06 0.21 0.06 0.22 0.07 -0.06 -0.03 -0.23 0.07 0.05 0.01 0.42 -0.33 0.09 0.01 0.16 0.10 0.73 1.00 0.71 0.67 0.66 0.39 0.24 0.39 0.35 0.07 0.39 0.30 SCRIS PMS - 0.19 -0.03 -0.04 0.24 0.07 0.24 0.12 -0.06 -0.03 -0.27 0.12 0.07 -0.02 0.37 -0.37 0.16 0.04 0.23 0.16 0.70 0.71 1.00 0.81 0.80 0.48 0.32 0.44 0.45 0.08 0.47 0.38 ORAL PMO - 0.20 0.00 -0.03 0.20 0.08 0.22 0.10 -0.06 -0.01 -0.23 0.10 0.04 0.02 0.35 -0.34 0.16 0.01 0.21 0.16 0.66 0.67 0.81 1.00 0.97 0.44 0.31 0.38 0.42 0.09 0.43 0.35 ORAL IO - 0.19 0.00 -0.03 0.20 0.07 0.22 0.09 -0.06 -0.01 -0.22 0.10 0.04 0.02 0.35 -0.34 0.16 0.00 0.21 0.15 0.66 0.66 0.80 0.97 1.00 0.43 0.30 0.37 0.41 0.08 0.42 0.35 PUNCTAI DIGITALE - 0.11 0.05 -0.09 0.18 0.12 0.15 0.01 -0.03 -0.03 -0.17 0.03 0.06 -0.03 0.18 -0.50 0.22 0.04 0.28 0.23 0.41 0.39 0.48 0.44 0.43 1.00 0.34 0.32 0.41 0.07 0.44 0.38 STATUS - 0.22 -0.11 -0.02 0.30 0.18 0.15 0.18 -0.03 -0.00 -0.31 0.20 0.11 0.02 0.32 -0.16 0.68 0.02 0.82 0.75 0.25 0.24 0.32 0.31 0.30 0.34 1.00 0.16 0.75 0.13 0.79 0.81 0.78 0.44 Medie - 0.17 -0.15 -0.05 0.15 0.05 0.11 0.05 -0.03 0.00 -0.14 0.04 0.13 0.02 0.30 -0.25 0.00 -0.02 0.00 0.00 0.36 0.39 0.44 0.38 0.37 0.32 0.16 1.00 0.44 0.06 0.42 0.34 0.45 -0.35 NOTA\_FINALA\_EA - 0.16 -0.19 -0.06 0.27 0.18 0.17 0.12 -0.07 -0.00 -0.27 0.13 0.12 0.13 0.39 -0.21 0.78 -0.09 0.64 0.62 0.35 0.35 0.45 0.42 0.41 0.41 0.75 0.44 1.00 0.15 NOTA FINALA EB - 0.02 -0.04 -0.03 0.04 0.01 0.01 -0.00 -0.07 0.00 -0.02 -0.01 -0.03 0.09 0.08 -0.05 0.14 0.15 0.11 0.10 0.06 0.07 0.08 0.09 0.08 0.07 0.13 0.06 0.15 1.00 0.12 0.12 0.16 0.03 NOTA FINALA EC - 0.16 -0.10 -0.04 0.38 0.16 0.20 0.21 -0.03 -0.00 -0.39 0.23 0.14 -0.01 0.38 -0.25 0.56 0.04 0.83 0.61 0.39 0.39 0.47 0.43 0.42 0.44 0.79 0.42 0.76 0.12 1.00 0.78 0.86 0.24 NOTA FINALA ED - 0.16 -0.14 -0.02 0.32 0.16 0.15 0.21 -0.03 0.01 -0.33 0.26 0.12 -0.01 0.34 -0.18 0.62 0.04 0.69 0.85 0.30 0.30 0.38 0.35 0.35 0.35 0.38 0.81 0.34 0.75 0.12 0.78 1.00 0.82 0.36 PCA1 - 0.22 -0.11 -0.04 0.45 0.18 0.29 0.26 -0.06 -0.01 -0.48 0.27 0.14 0.01 0.51 -0.37 0.03 0.16 PCA2 - -0.06 -0.15 0.15 0.09 0.08 -0.13 0.23 0.07 0.06 -0.14 0.22 0.05 0.02 -0.12 0.35 0.52 0.00 0.51 0.56 -0.54 -0.55 -0.53 -0.55 -0.55 -0.25 0.44 -0.35 0.24 0.03 STATUS\_ED candidat

1.00

- 0.75

0.50

- 0.25

0.00

-0.25

-0.50

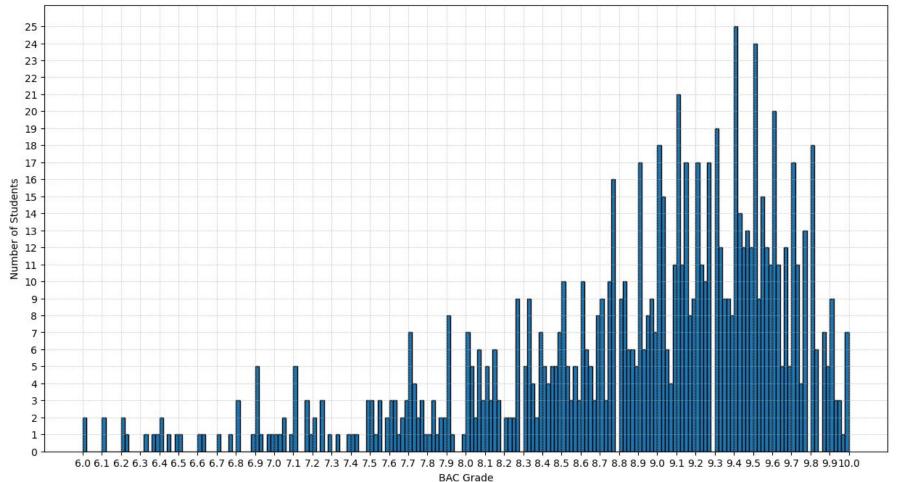
-0.75

## **Dataset Admiteri - Vizualizare**

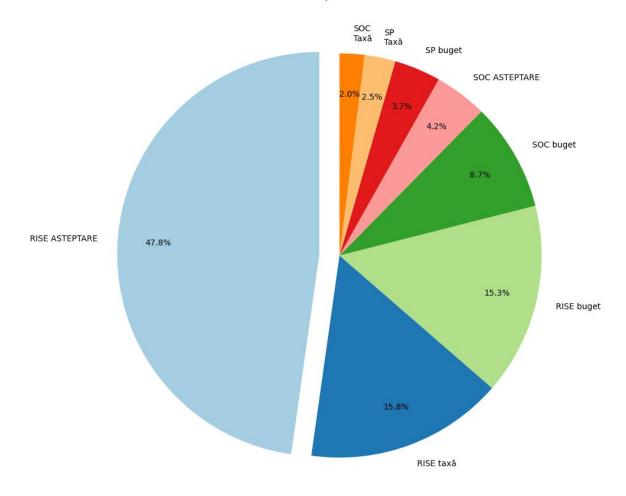
Această analiză își propune să evidențieze:

- 1. Distribuția notelor de admitere la diverse programe universitare din România, precum UNIBUC (CTI, Mate), UPB (ACS), UMFCD etc.
- 2. Corelația între performanțele la Bacalaureat și notele obținute la anumite examene specifice, precum examenul de Română (Drept).
- 3. Popularitatea specializărilor și statutul admiterii în funcție de datele disponibile pentru fiecare instituție.

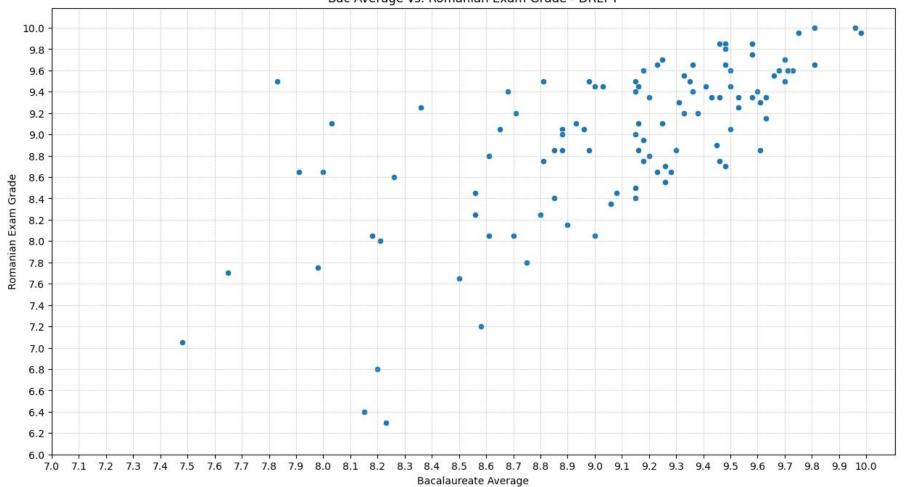
Distribution of BAC Grades - UPB ACS



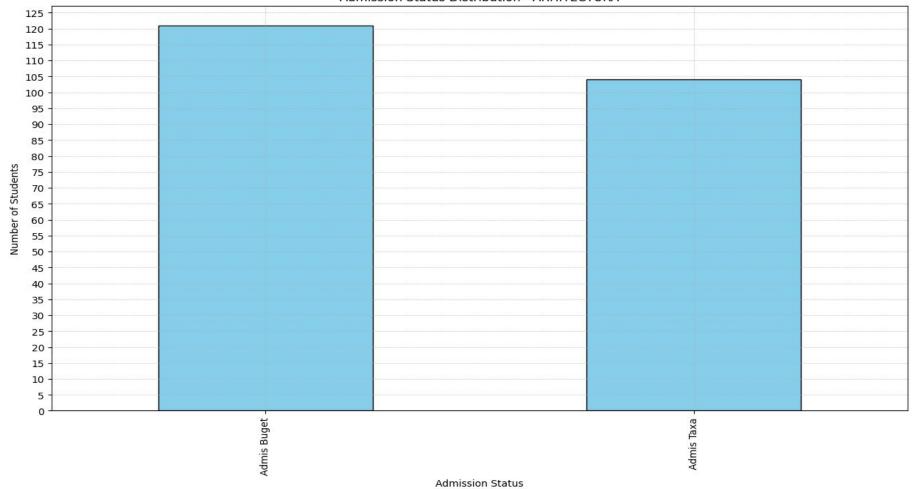
#### Distribution of Specializations - SNSPA



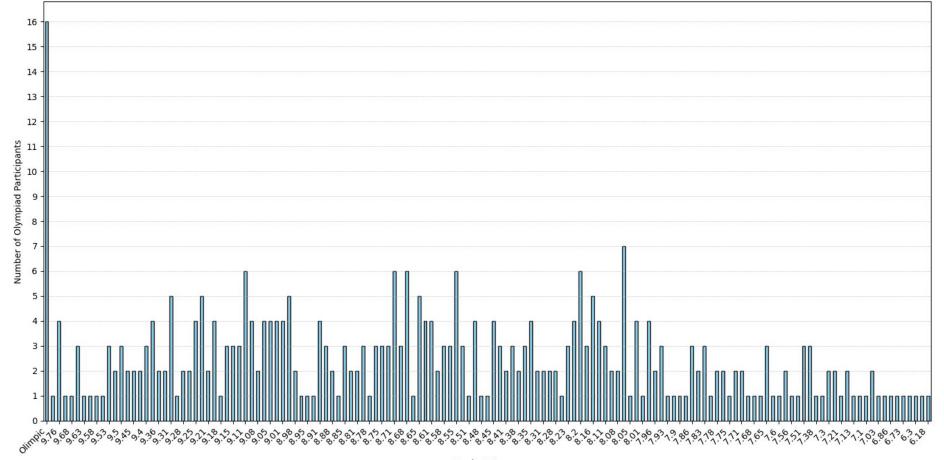
Bac Average vs. Romanian Exam Grade - DREPT



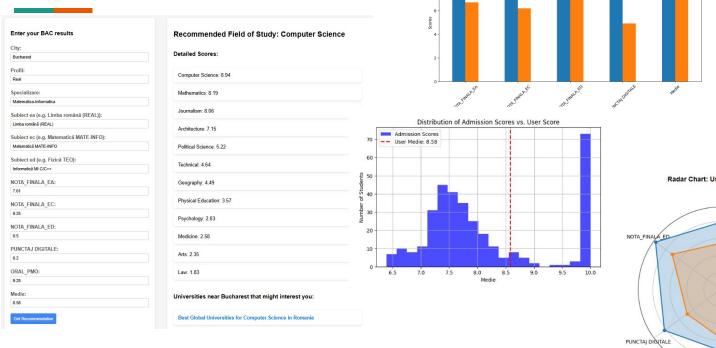
#### Admission Status Distribution - ARHITECTURA



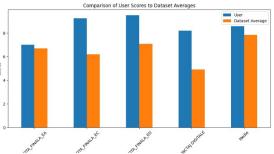
#### Olympiad Participation - GEOGRAFIE



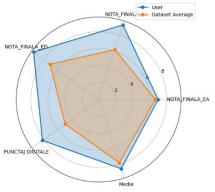
# Demo Aplicație Web



#### Relevant Charts for your results:



#### Radar Chart: User vs. Dataset Average



# Referințe

- https://docs.jupyter.org/en/latest/
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- https://realpython.com/k-means-clustering-python/
- https://flask.palletsprojects.com/en/stable/tutorial/
- https://serpapi.com/search-api
- http://ossasepia.com/bac-data/