

HUSHHUSH RECRUITER

Attracting, developing, and retaining top corporate talent.



TEAM MEMBERS

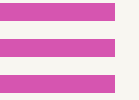
SAKSHI MANJREKAR
ISHWARI THAKUR
SIMRAN AMESAR





PROBLEM STATEMENT

hushHush Recruiter requires a talent selection system that analyzes online technical activity and manages coding assessments end-to-end.

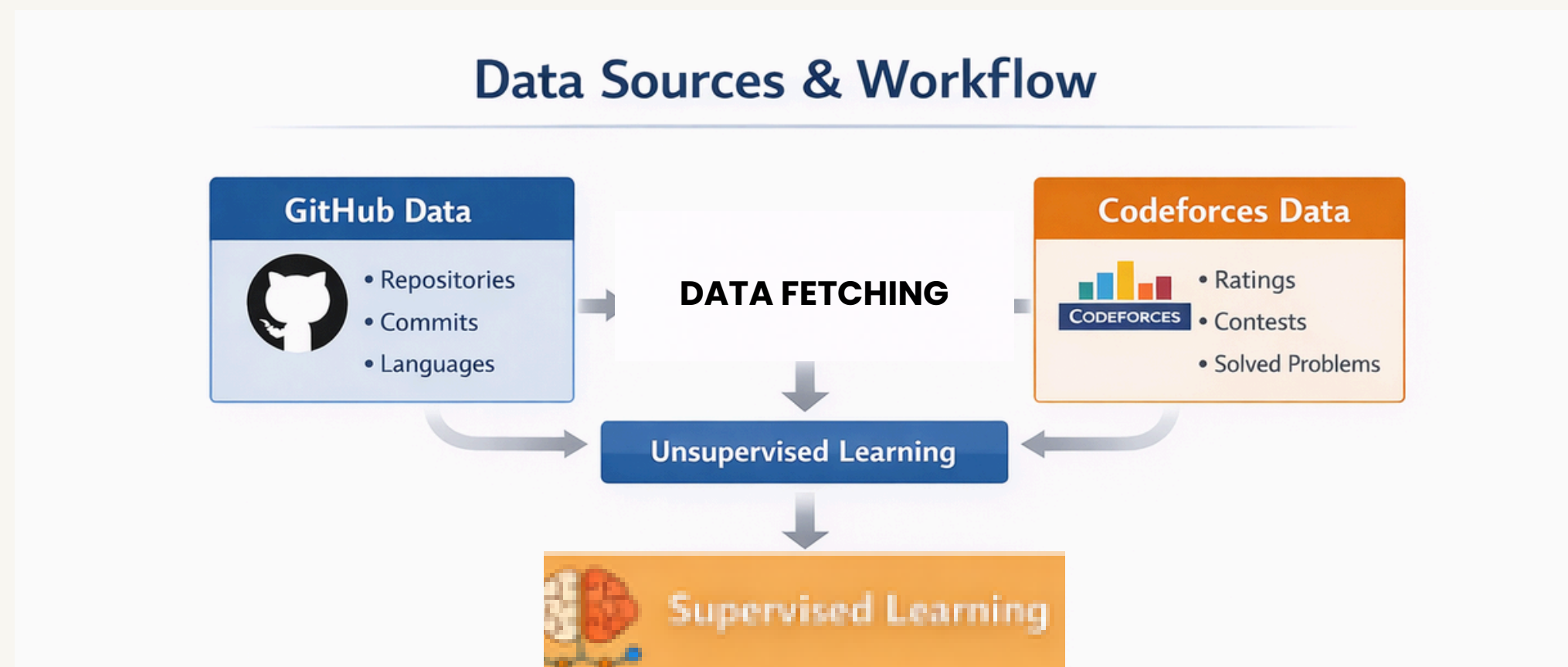


DATA SOURCES

GITHUB

CODEFORCES

DATASET OVERVIEW



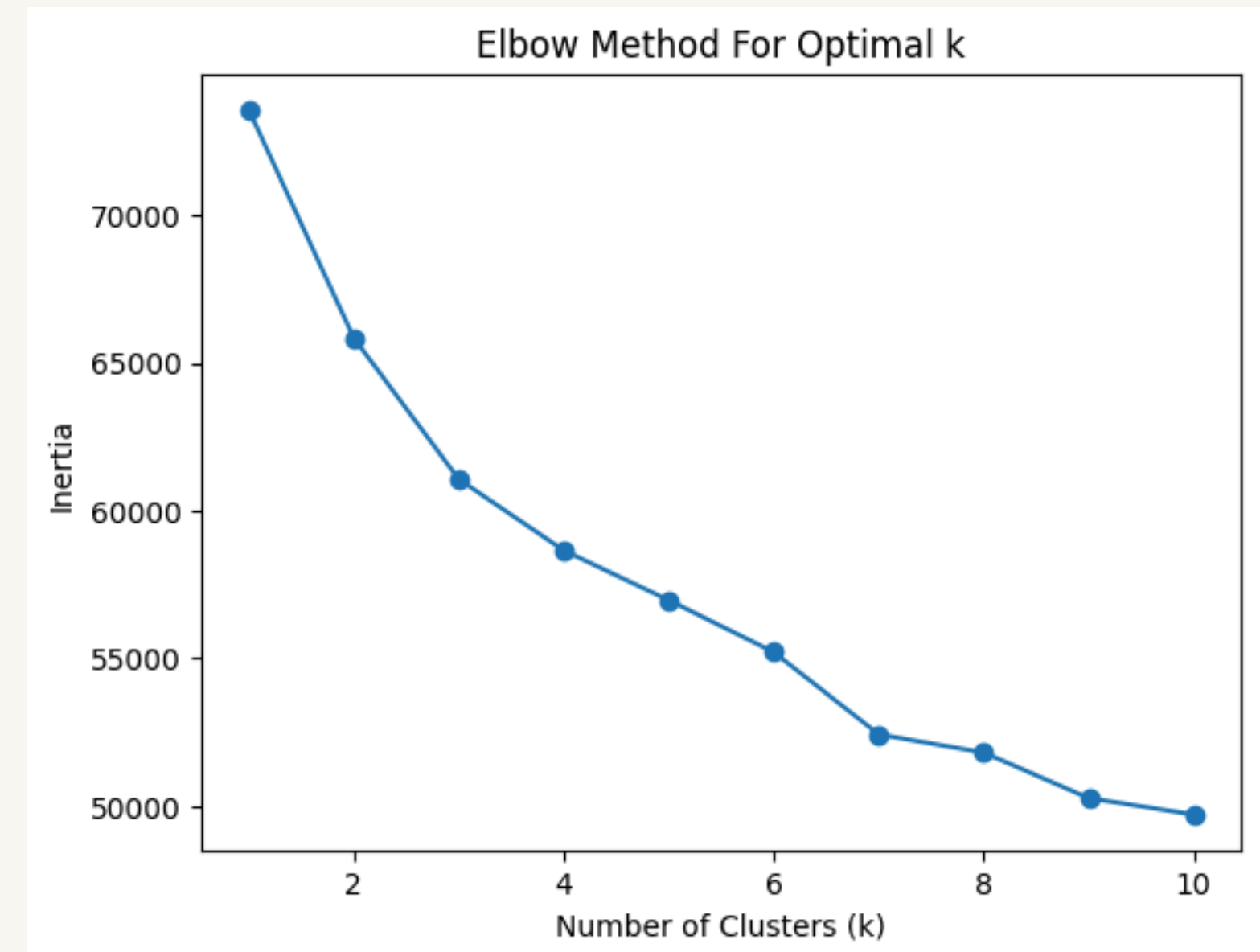
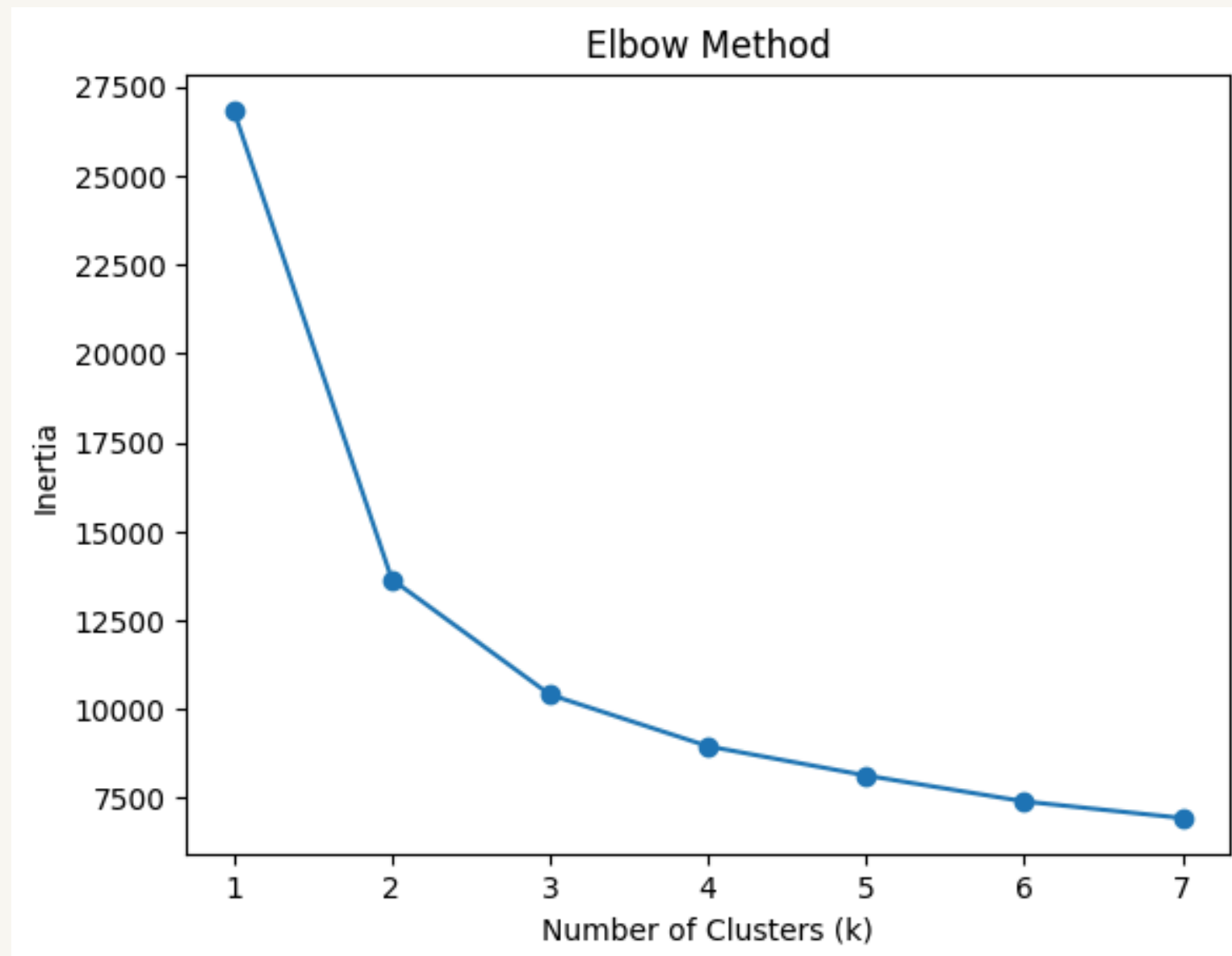


GITHUB

ELBOW METHOD

1. RANDOM FOREST

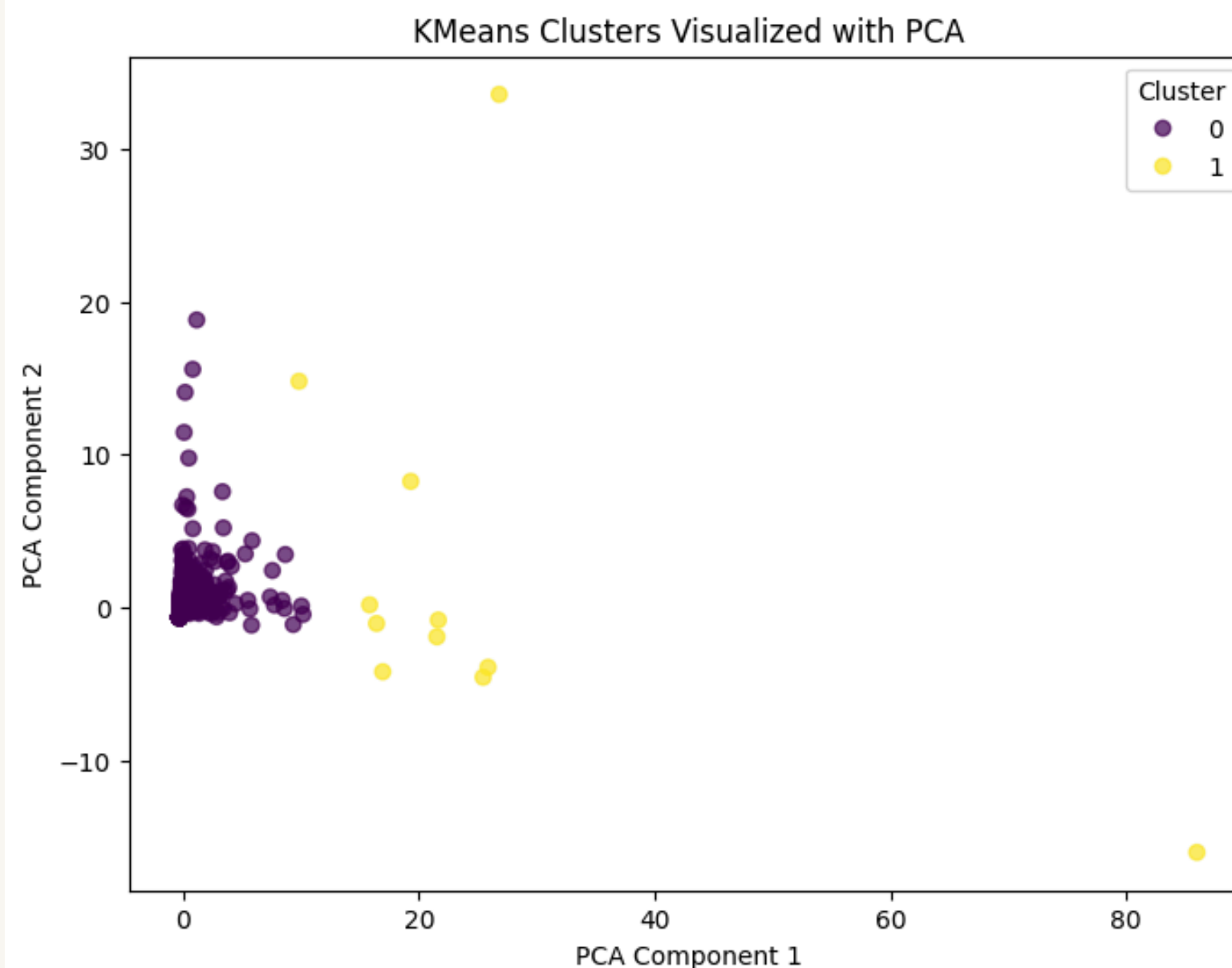
2. DECISION TREE



CLUSTERS USING PCA

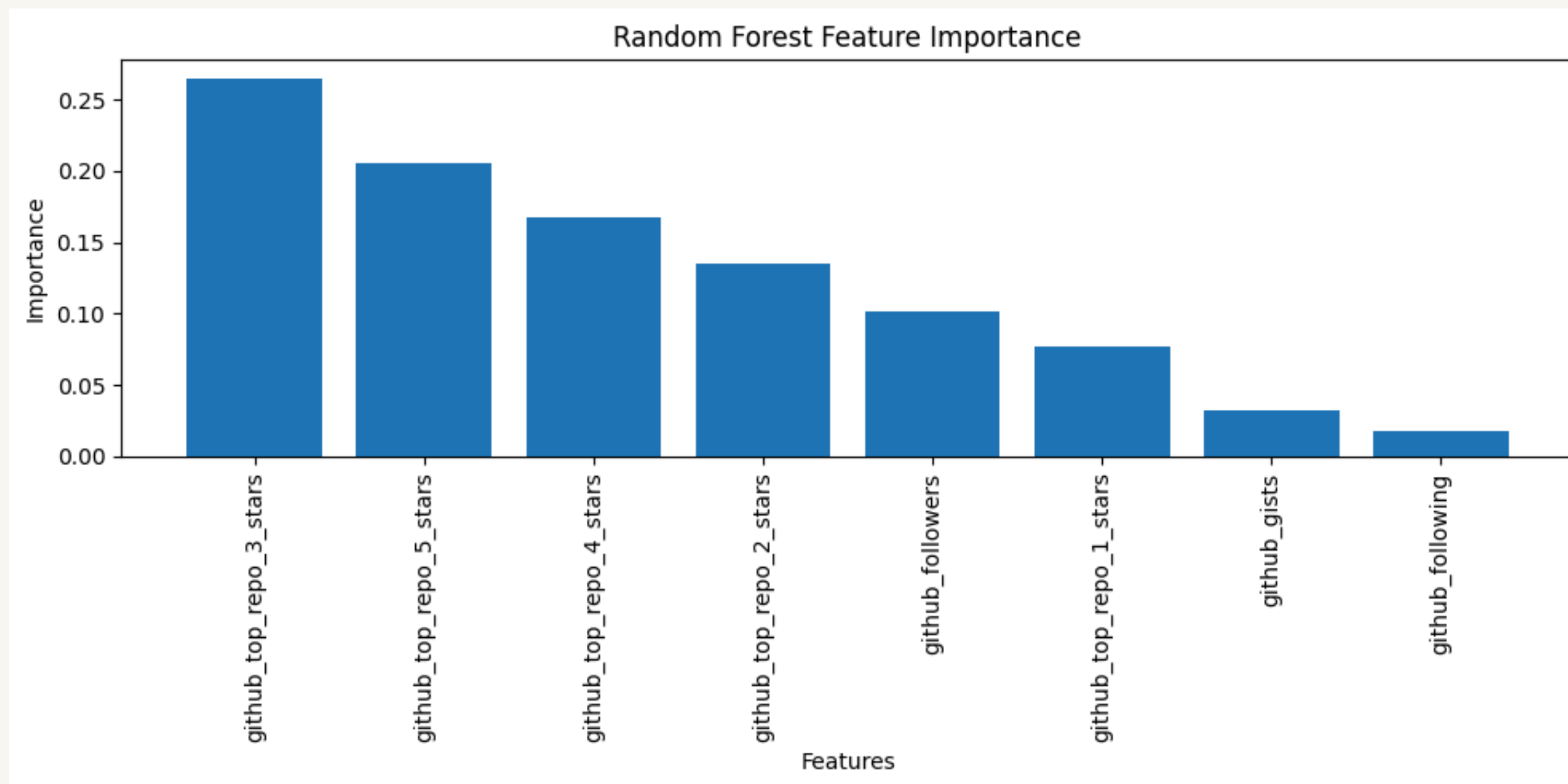
1. RANDOM FOREST

2. DECISION TREE



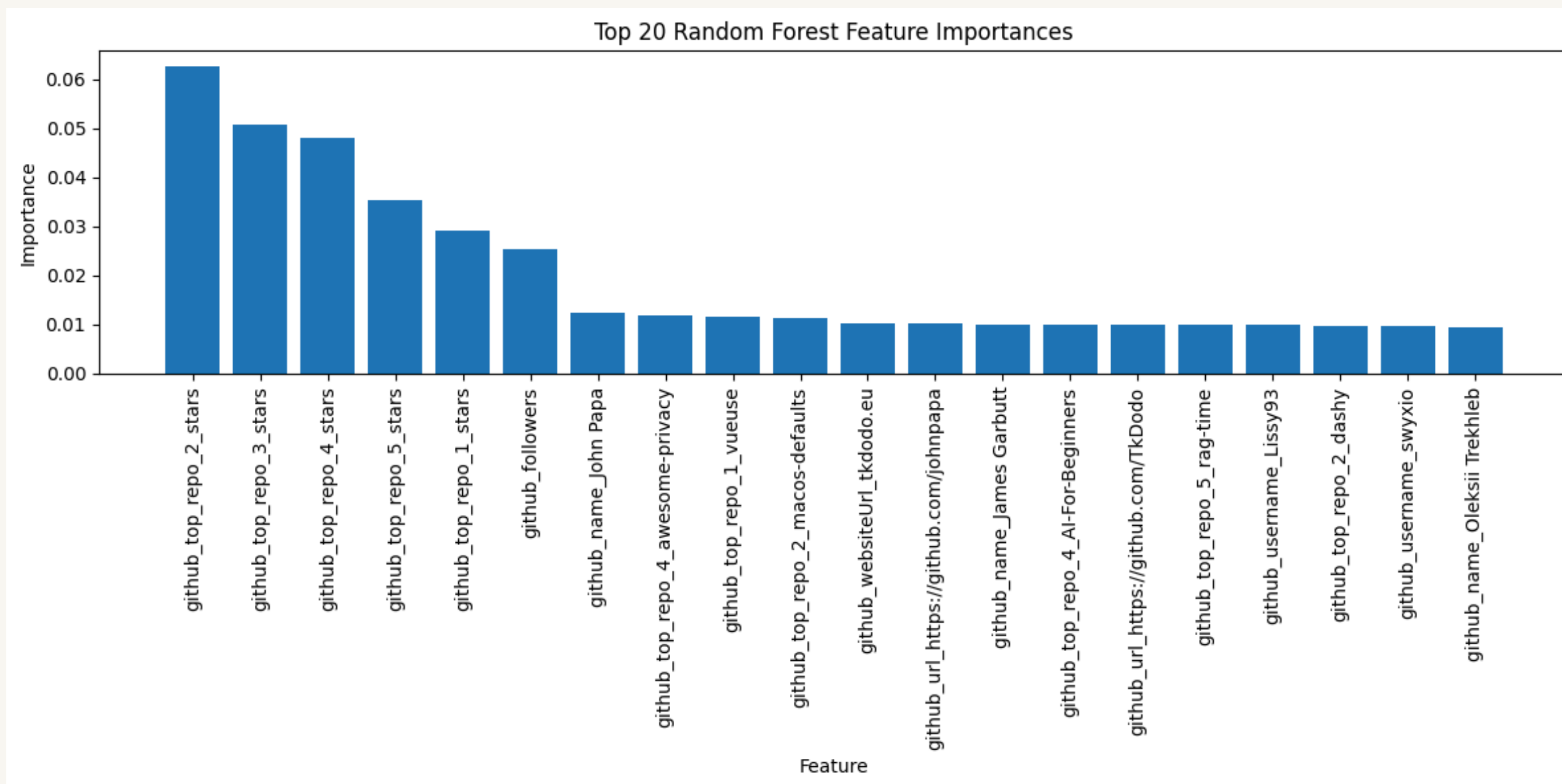
FEATURE IMPORTANCE

1. RANDOM FOREST



FEATURE IMPORTANCE

2. DECISION TREE



1. RANDOM FOREST

Top 10 GitHub Technical Users:

	user_name	github_username	github_followers	...	github_top_repo_4_stars	github_top_repo_5_stars	technical_score
2263	Yann Bertrand	yannbertrand	201.0	...	42.0	16.0	396164.0
1636	John Papa	johnpapa	15368.0	...	45122.0	33665.0	378609.0
2939	Evan You	yyx990803	106605.0	...	4487.0	4050.0	351530.0
2790	Oleksii Trekhleb	trekhleb	17724.0	...	6126.0	2266.0	263360.0
1941	Chris Noring	softchris	952.0	...	2789.0	1356.0	200427.0
2485	Dominik D	TkDodo	3276.0	...	5259.0	5214.0	161891.0
1580	Sebastien Lorber	slorber	2036.0	...	5387.0	2145.0	154566.0
738	Nitya Narasimhan, Ph.D	nitya	671.0	...	16740.0	772.0	151021.0
1246	Ryan Michael	kerinin	30.0	...	34.0	24.0	144917.0
3212	Dmitriy Kovalenko	dmtrKovalenko	1195.0	...	2750.0	1705.0	117281.0

2. DECISION TREE

Top 10 Technical Users Based on GitHub Metrics:

	github_username	github_login	github_name	github_location	...	github_top_repo_5	github_top_repo_5_stars	Cluster	technical_score
2263	yannbertrand	yannbertrand	Yann Bertrand	France	...	avatars-io	16.0	1	396164.0
1636	johnpapa	johnpapa	John Papa	Orlando, FL	...	Data-Science-For-Beginners	33665.0	1	378609.0
2939	yyx990803	yyx990803	Evan You	Singapore	...	proposal-signals	4050.0	1	351530.0
2790	trekhleb	trekhleb	Oleksii Trekhleb	San Francisco Bay Area	...	nano-neuron	2266.0	1	263360.0
1941	softchris	softchris	chris	London	...	react-book	1356.0	1	200427.0
2485	TkDodo	TkDodo	Dominik Dorfmeister	Vienna, Austria	...	react-18	5214.0	1	161891.0
1580	slorber	slorber	Sébastien Lorber	Paris	...	react-async	2145.0	1	154566.0
738	nitya	nitya	Nitya Narasimhan	NY	...	rag-time	772.0	1	151021.0
1246	kerinin	kerinin	Ryan Michael	Durham NC	...	arduino-thermostat	24.0	0	144917.0
3212	dmtrKovalenko	dmtrKovalenko	Dmitriy Kovalenko	In your node_modules	...	toolpad	1705.0	0	117281.0

MODEL PERFORMANCE

MODEL	ACCURACY
Random Forest	99%
Decision Tree	99%

WHY RANDOM FOREST?

- 1.It uses more features (all numeric technical metrics).
- 2.It is an ensemble – generally higher and more stable accuracy.

BEST MODEL: RANDOM FOREST

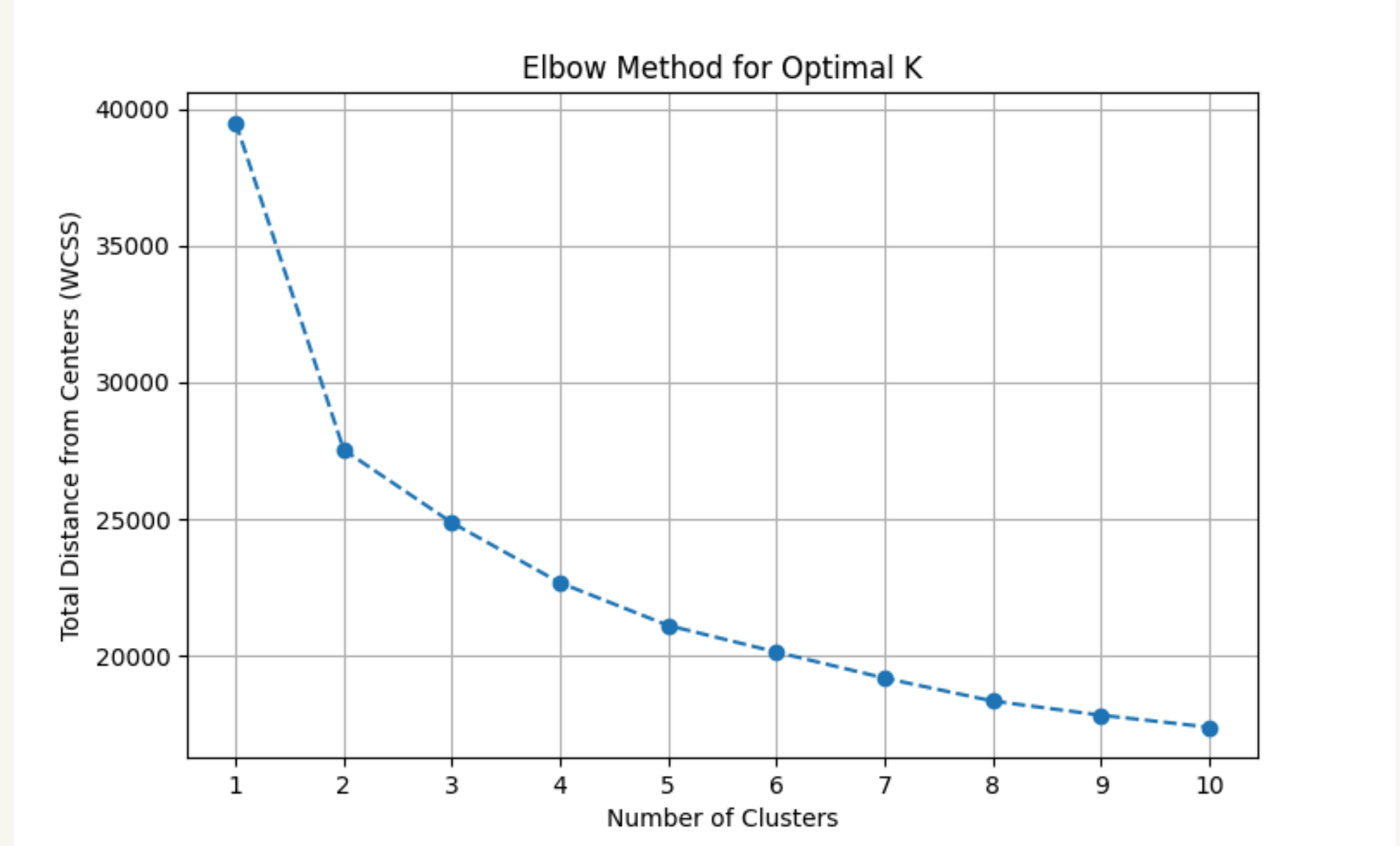
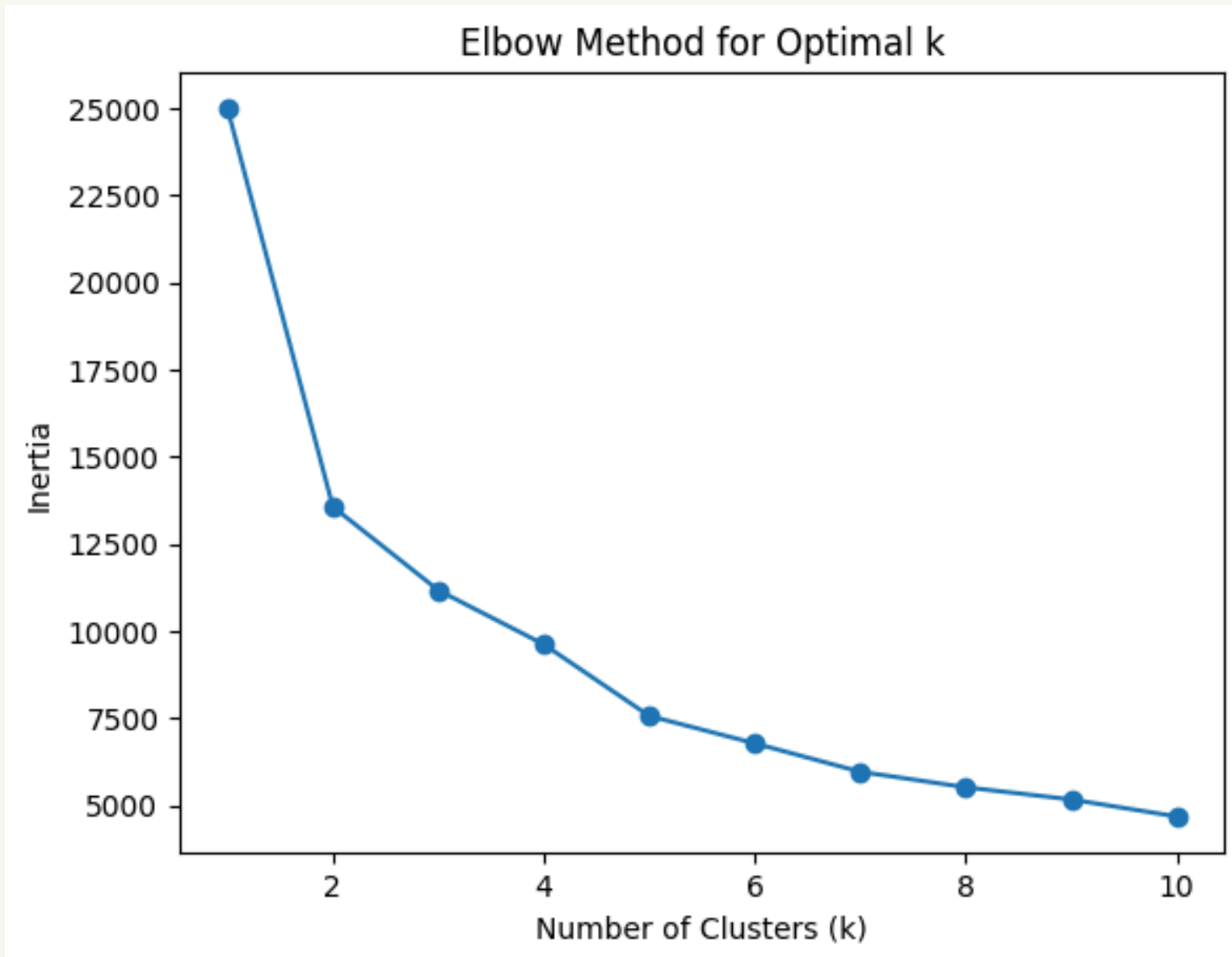


CODEFORCES

ELBOW METHOD

1. RANDOM FOREST

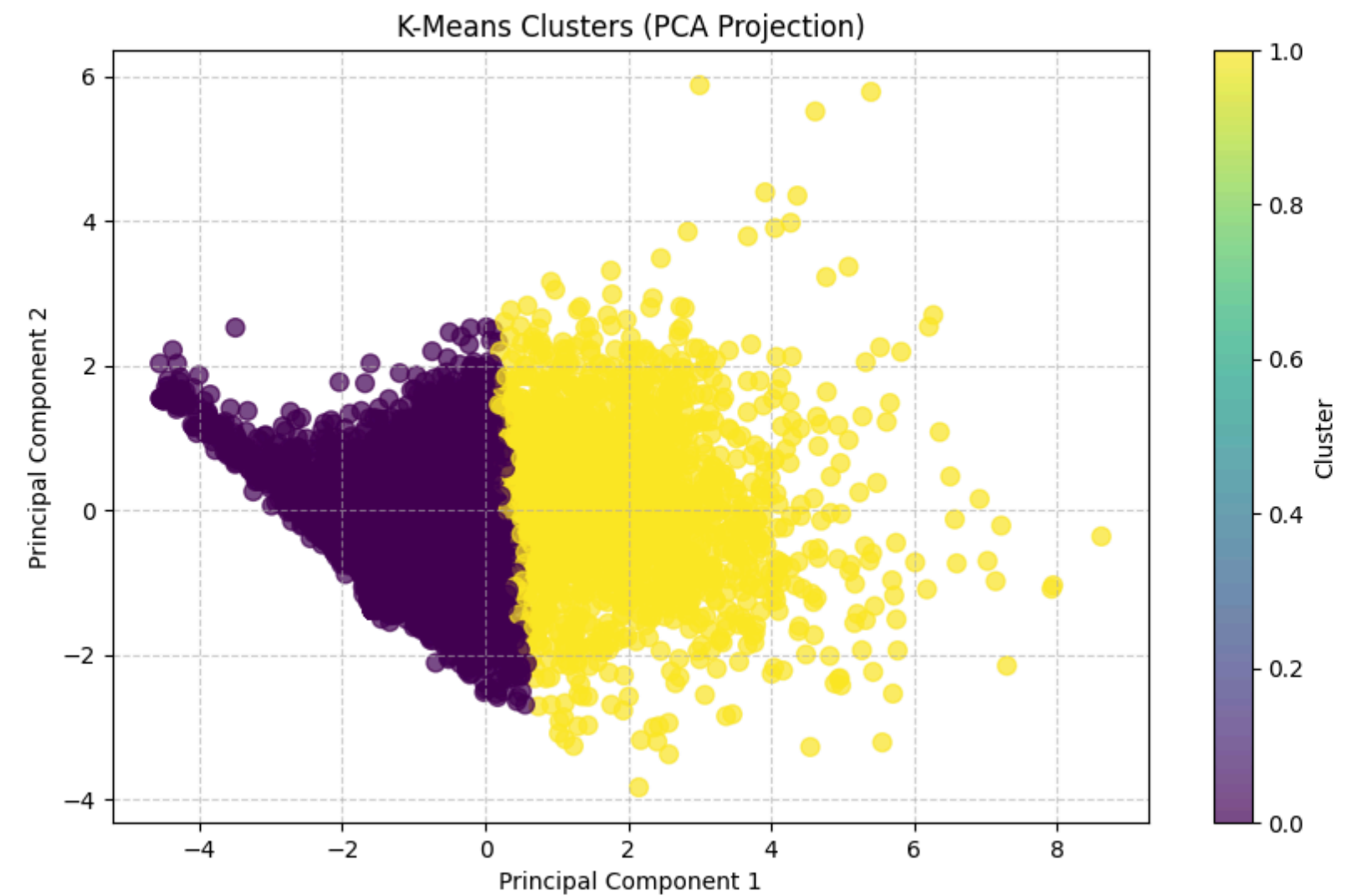
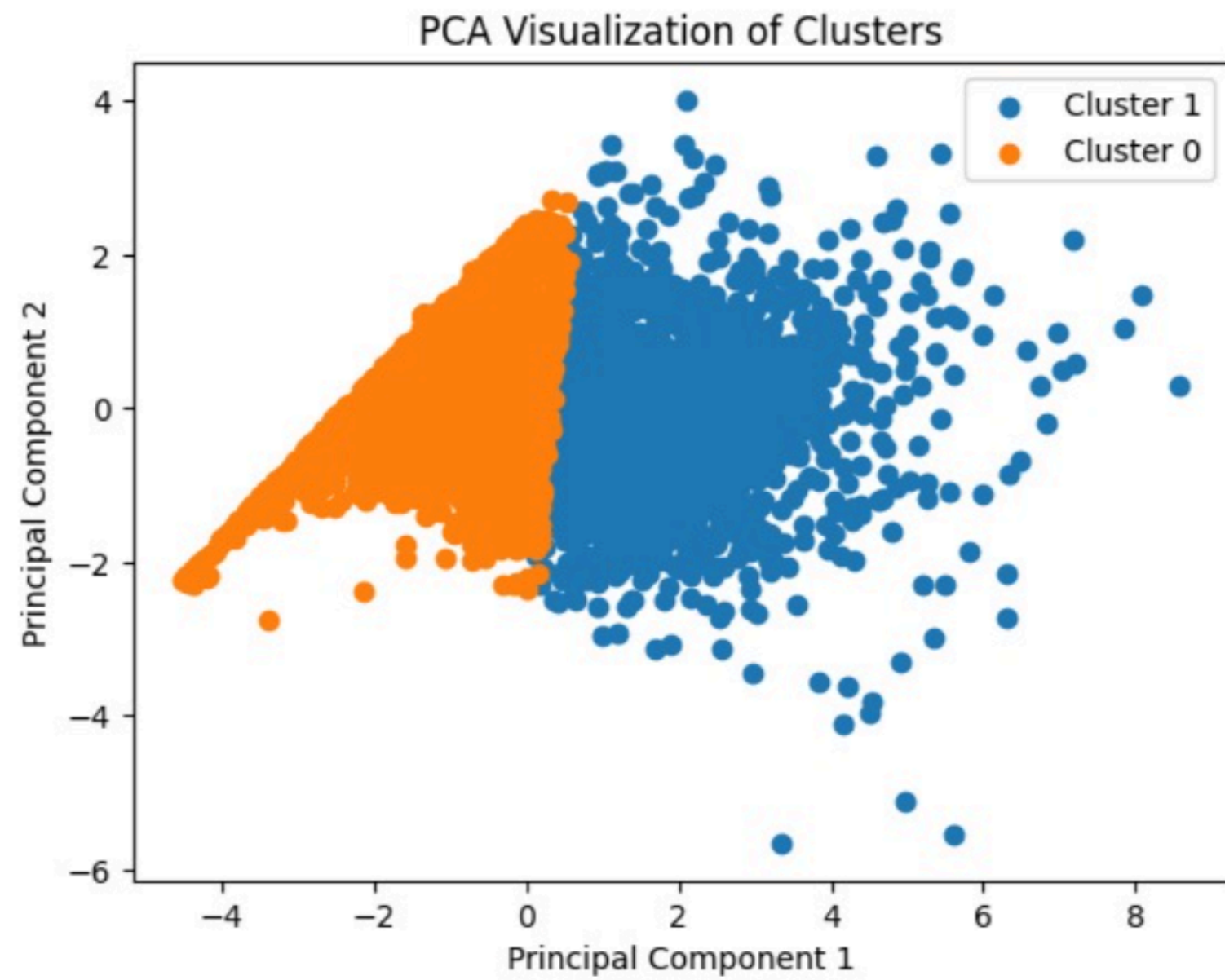
2. GRADIENT BOOSTING



CLUSTERS USING PCA

1. RANDOM FOREST

2. GRADIENT BOOSTING



1. RANDOM FOREST

	username	rating	total_solved	hard_solved	contests
3936	Benq	3792.000000	840	638	166
2851	jiangly	3710.000000	1221	643	191
2134	ecnerwala	3666.000000	1084	710	213
3912	Kevin114514	3658.000000	532	414	103
65	VivaciousAubergine	3533.000000	516	370	64

2. GRADIENT BOOSTING

	username	rating	max_rating	total_solved	hard_solved	contests
3936	Benq	3792.0	3833.0	840	638	166
2851	jiangly	3710.0	4039.0	1221	643	191
2134	ecnerwala	3666.0	3741.0	1084	710	213
3912	Kevin114514	3658.0	3811.0	532	414	103
65	VivaciousAubergine	3533.0	3533.0	516	370	64

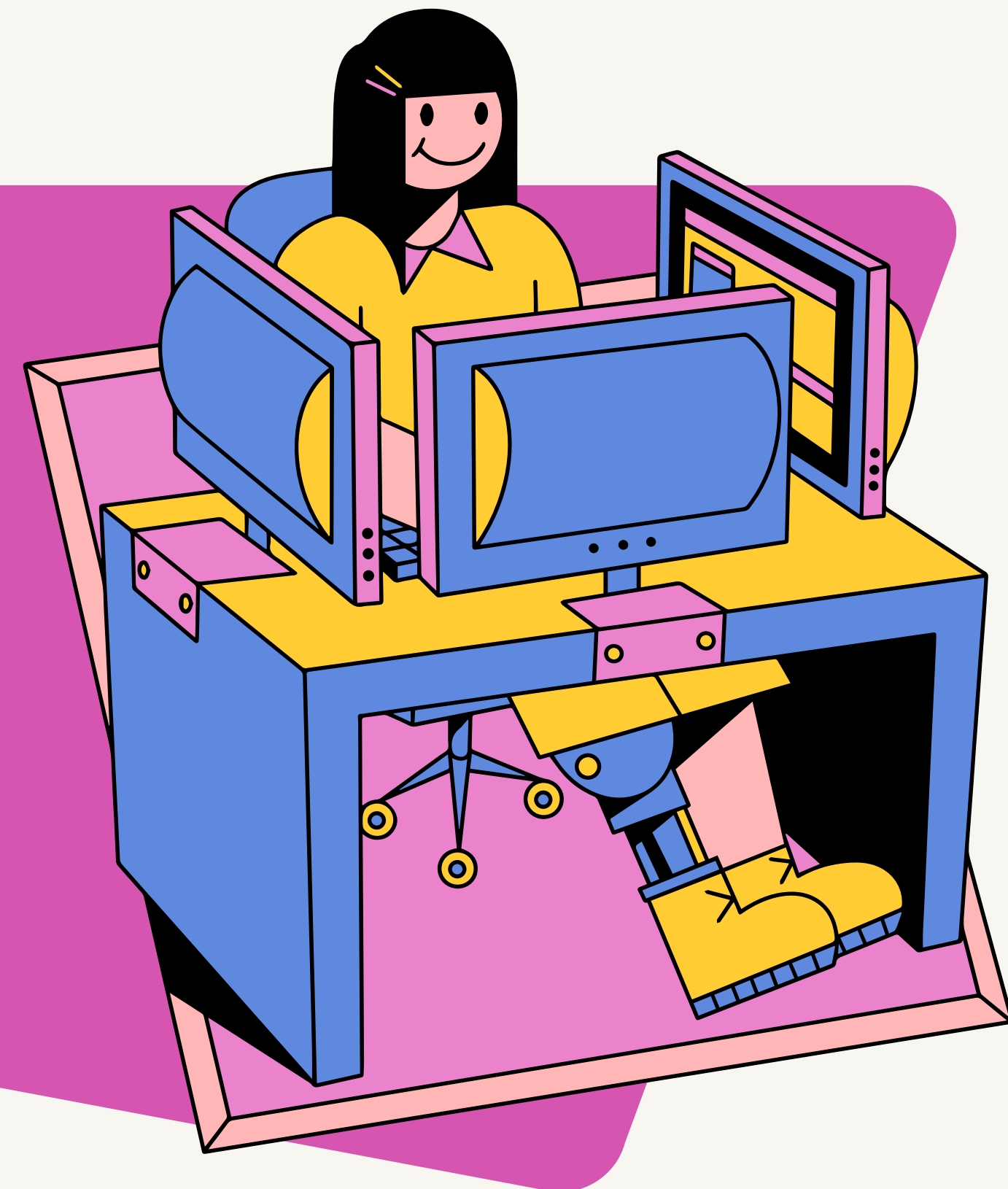
MODEL PERFORMANCE

MODEL	ACCURACY
Random Forest	96.70%
Gradient Boosting	97.30%

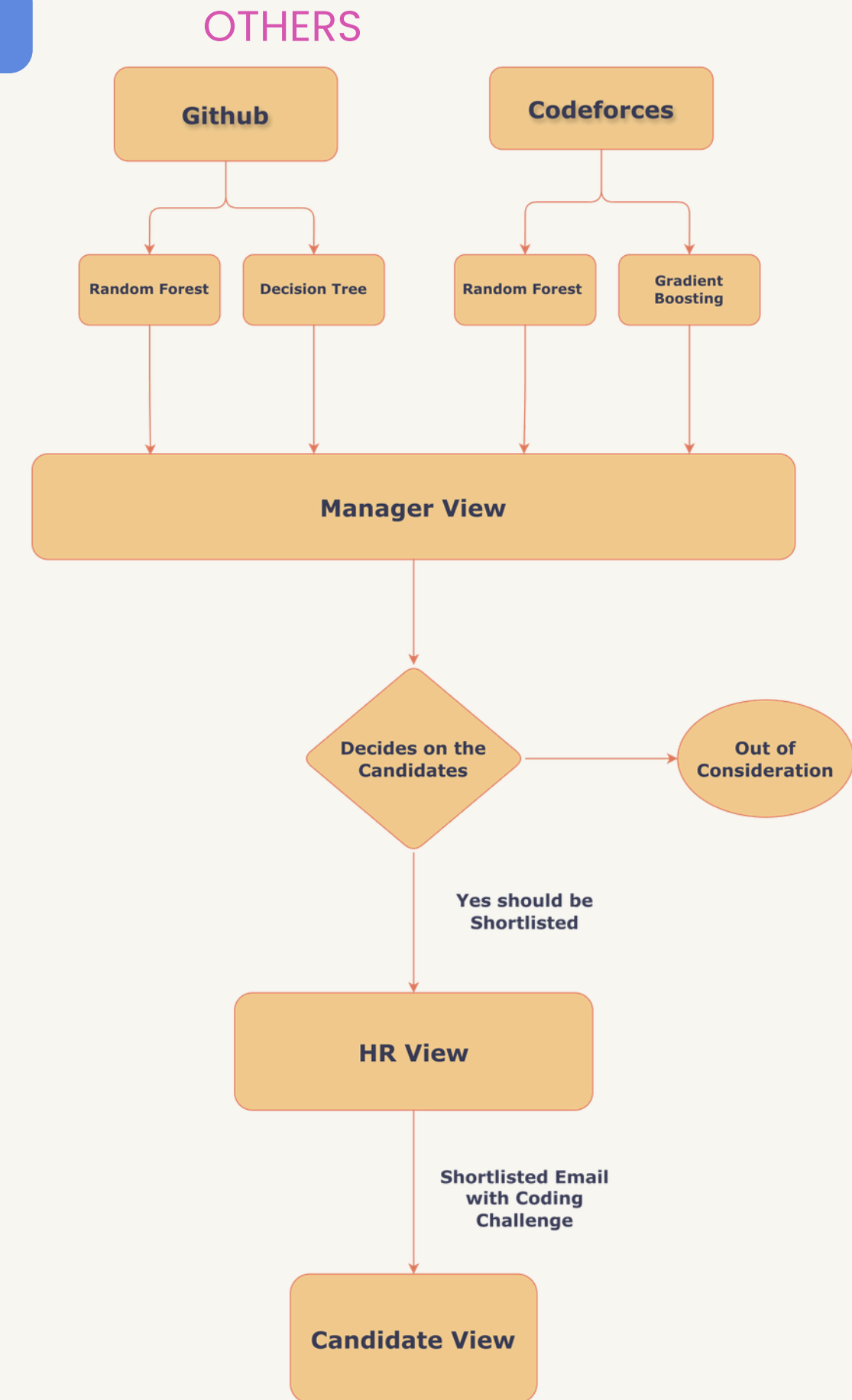
WHY GRADIENT BOOSTING?

1. It captures complex feature interactions (tags, languages, numeric metrics) better.
2. Outperforms Random Forest in distinguishing subtle differences among users.

BEST MODEL: GRADIENT BOOSTING

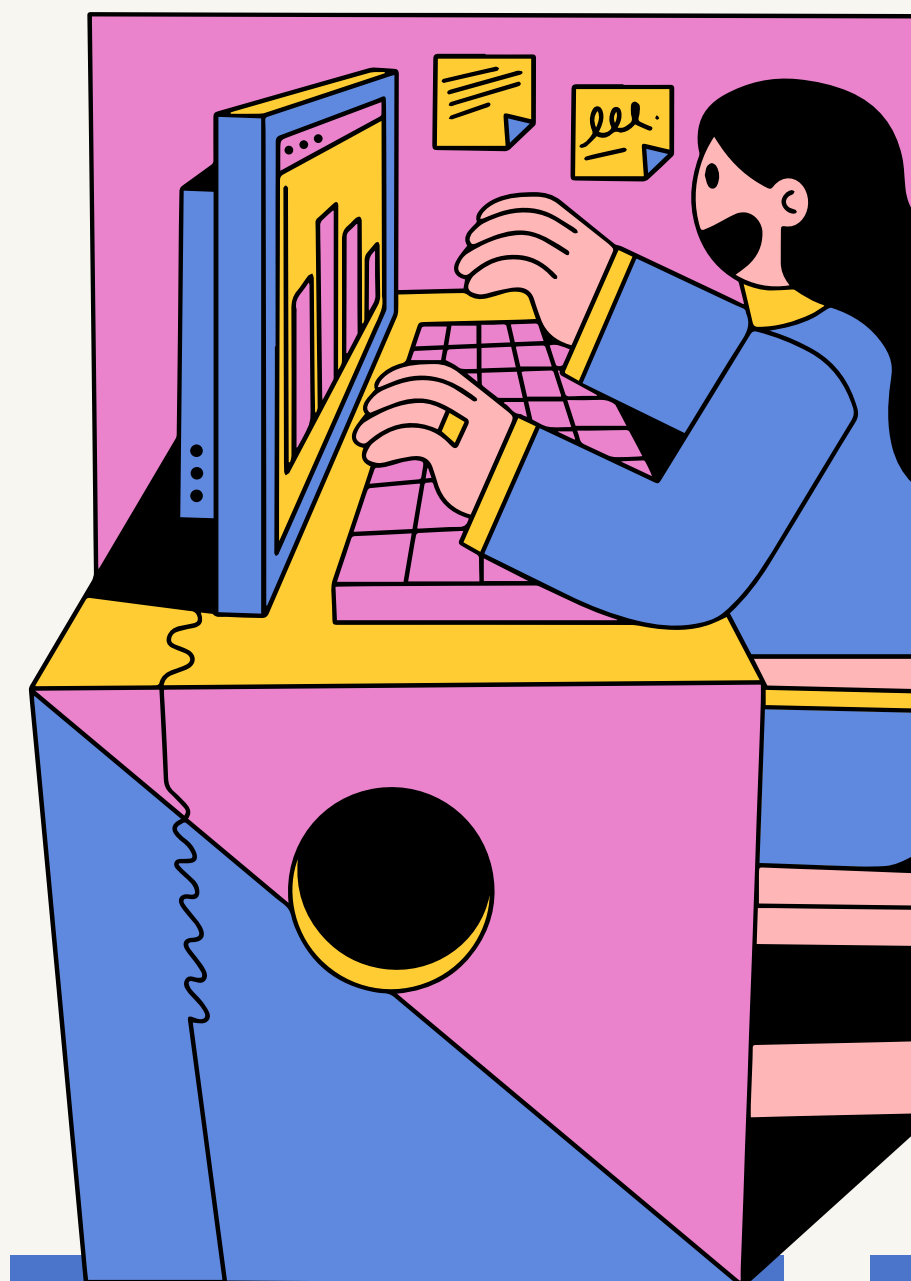


UI DEMO AND ARCHITECHTURE



FURTHER IMPROVEMENTS

- Own Code Challenges
- Location Filter for Candidates
- Skill Sorting according to requirement
- User Persona Login System





THANK YOU