URBAN SYSTEMS

The Final Competition Symposium

Network Analytics and Machine Learning for Employee Retention and Productivity Prediction

Presented by: Lalith Nandakumar & Binh Nguyen

Agenda

- Introduction
- Key Approaches
- Key Questions
- Insights
- Conclusion

Introduction

- Employee retention and productivity are key drivers of organizational success
- Understanding workforce dynamics requires analyzing performance, engagement, and attrition factors
- This project integrates graph based analytics and machine learning to uncover insights into employee relationships and performance

Key Approaches

- Graph-based features:
  - PageRank
  - Betweenness Centrality
  - Louvain Community Detection
  - Dijkstra’s Shortest Path Algorithm
- Machine learning models:
  - Logistic Regression
  - Random Forest
  - XGBoost

Graph-based features

PageRank

- Runs PageRank to measure the importance or influence of each employee in the network.
- Stores the result in the page_rank property of each Employee node.
- Employees with higher PageRank are more influential, meaning they are frequently connected to key employees (e.g., senior employees, mentors).

Betweenness Centrality

- Computes Betweenness Centrality, which measures how often an employee lies on the shortest path between other employees.
- Stores the score in betweenness_centrality.
- Employees with high Betweenness Centrality are important connectors.
- If removed, they could disrupt communication between different employee groups.

Graph-based features

Louvain Community Detection

- Detects clusters (communities) of employees who are more closely connected to each other than to the rest of the network.
- Assigns a community ID to each employee and stores it in community.
- Employees in the same community are more likely to collaborate.
- Community structure can reveal functional groups, isolated teams, or hidden bottlenecks.

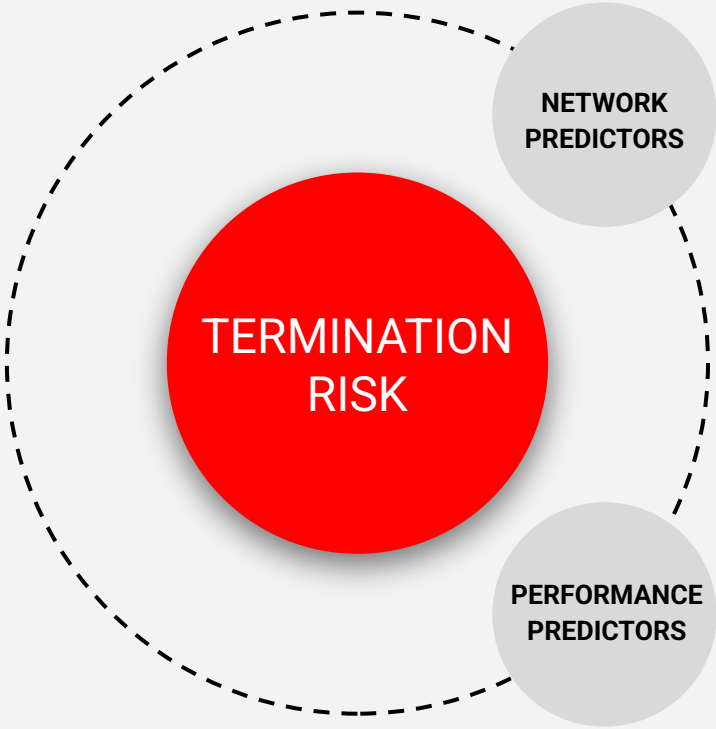
Dijkstra’s Shortest Path Algorithm

- Uses Dijkstra’s Shortest Path Algorithm to calculate the shortest path from each employee to their coach.
- Saves the result in distance_to_coach.
- Employees with shorter distances to their coach have stronger mentorship and may receive better guidance.
- Employees far from a coach might struggle with learning, performance, or retention.

Key Questions

1. Can we predict which employees are at risk of termination based on network-related and performance-related predictors?
2. How do an employee's network position and their work engagement influence their retention and productivity?
3. What factors contribute to project success or failure?
4. How does partner status moderate the relationship between an employee's centrality in the coaching network and their overall billable performance?

1. **Can we predict which employees are at risk of termination based on network-related and performance-related predictors?**



- 'page_rank'
- 'betweenness'
- 'distance_to_coach'

- 'target_ratio'
- 'rate'

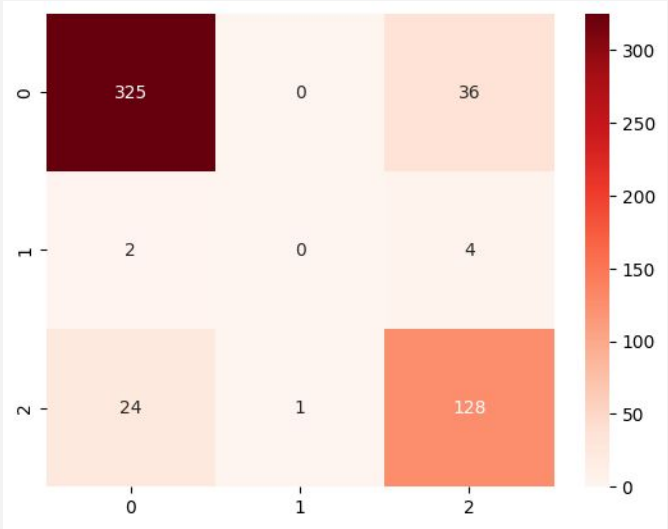
Random Forest vs. XGBoost

Random Forest Classifier Report:

	precision	recall	f1-score	support
0	0.91	0.93	0.92	355
1	0.00	0.00	0.00	6
2	0.81	0.81	0.81	159
accuracy			0.88	520
macro avg	0.57	0.58	0.58	520
weighted avg	0.87	0.88	0.88	520

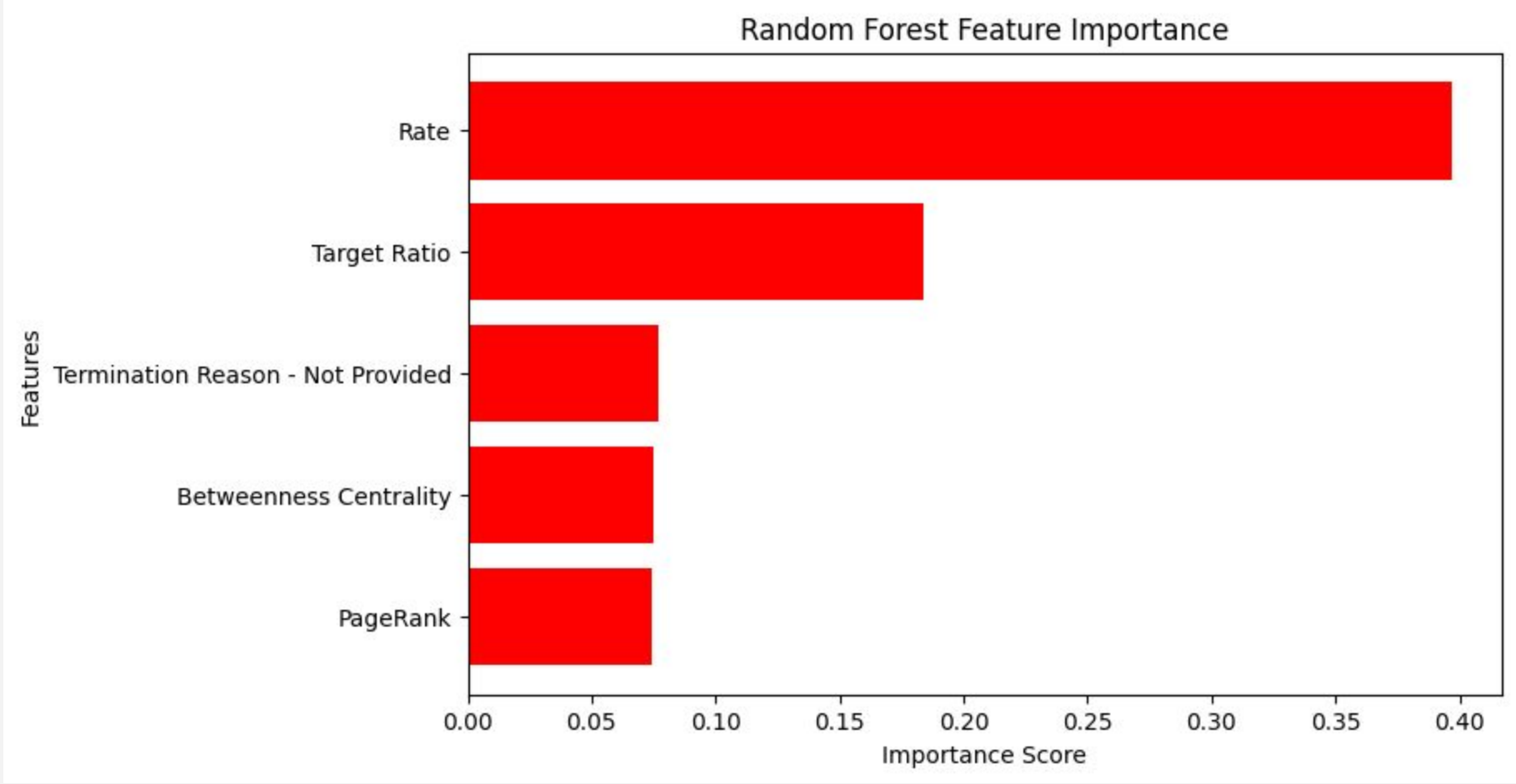
XGBoost Classifier Report:

	precision	recall	f1-score	support
0	0.91	0.89	0.90	361
1	0.00	0.00	0.00	6
2	0.73	0.80	0.77	153
accuracy			0.85	520



Status

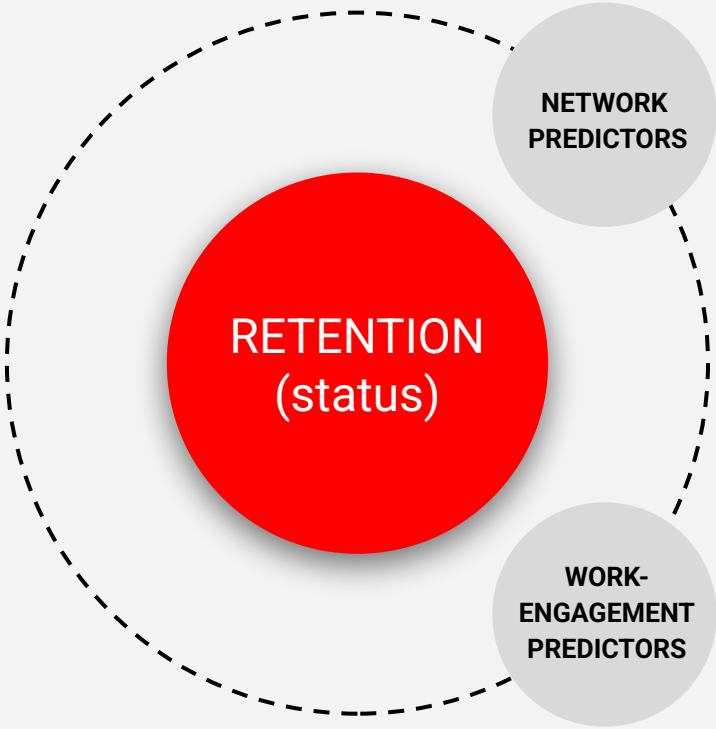
0: Terminated
1: Inactive
2: Active


Random Forest Feature Importance
Rate
Target Ratio
Termination Reason - Not Provided
Betweenness Centrality
PageRank
Features
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Importance Score

Insights

- Rate and Target_ratio are the strongest predictors of termination.
- Employees with weaker organizational connections are at higher risk of termination.
- Network position plays a secondary role but still influences retention.
- Workplace relationships, while less predictive than performance metrics, contribute to employee stability.

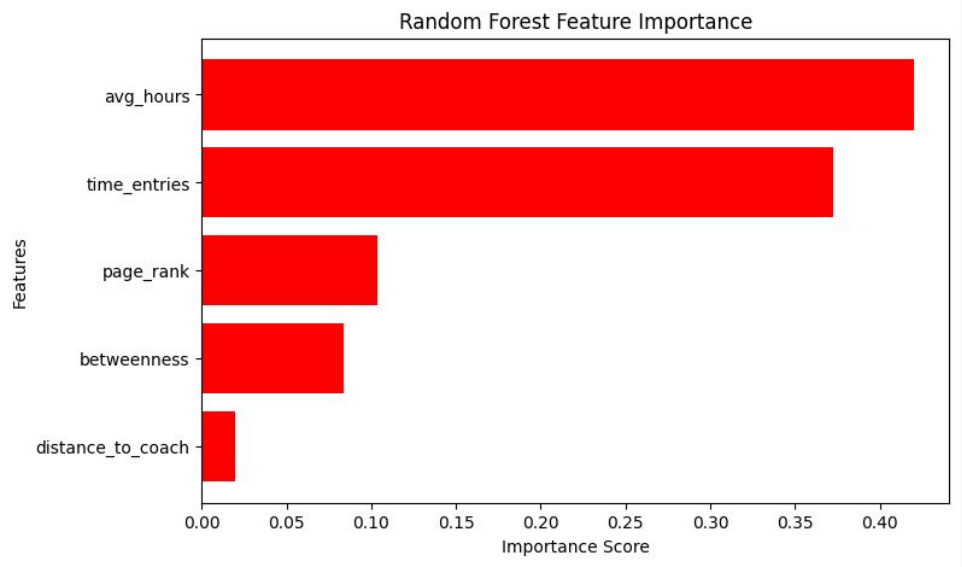
2. How do an employee's network position and their work engagement influence their retention and productivity?



- 'page_rank'
- 'betweenness'
- 'distance_to_coach'

- 'time_entries'
- 'avg_hours'

Random Forest, XGBoost, and Logistic Regression Comparison



Status

0: Terminated | 1: Inactive | 2: Active

Random Forest Classifier Report:

	precision	recall	f1-score	support
0	0.82	0.92	0.87	361
1	0.00	0.00	0.00	6
2	0.72	0.54	0.62	153
accuracy			**0.80**	520
macro avg	0.51	0.49	0.50	520
weighted avg	0.78	0.80	0.79	520

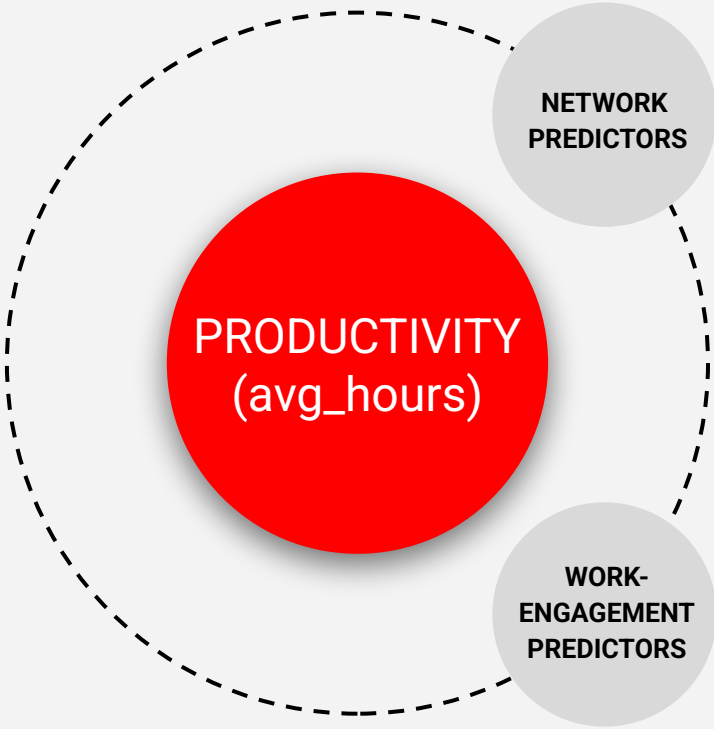
XGBoost Classifier Report:

	precision	recall	f1-score	support
0	0.82	0.90	0.86	355
1	0.00	0.00	0.00	6
2	0.71	0.59	0.65	159
accuracy			**0.80**	520
macro avg	0.51	0.50	0.50	520
weighted avg	0.78	0.80	0.79	520

Logistic Regression Report:

	precision	recall	f1-score	support
0	0.65	0.85	0.73	183
1	0.00	0.00	0.00	7
2	0.69	0.47	0.56	154
accuracy			**0.66**	344
macro avg	0.45	0.44	0.43	344
weighted avg	0.65	0.66	0.64	344

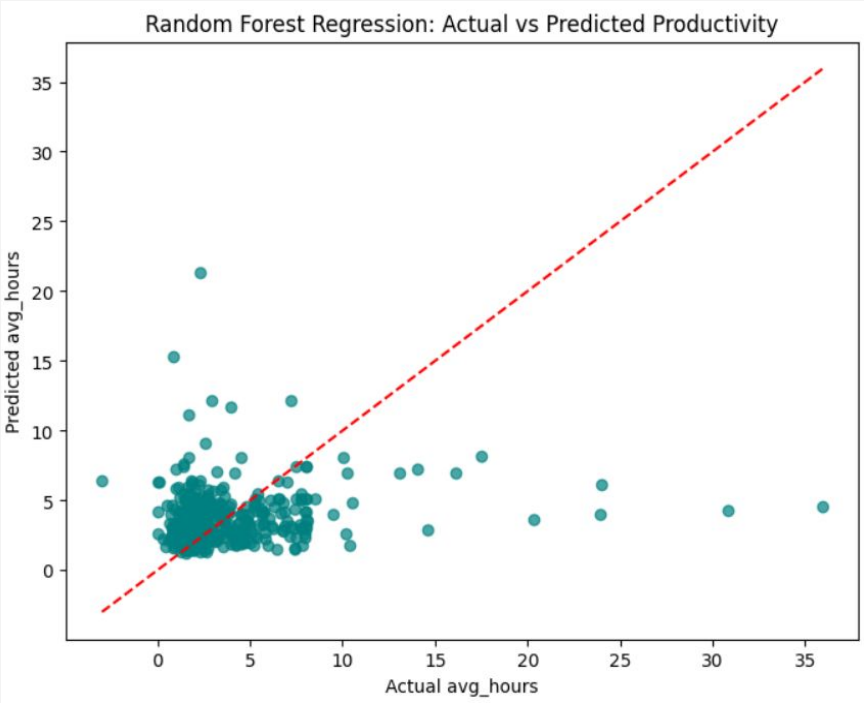
Predicting Productivity

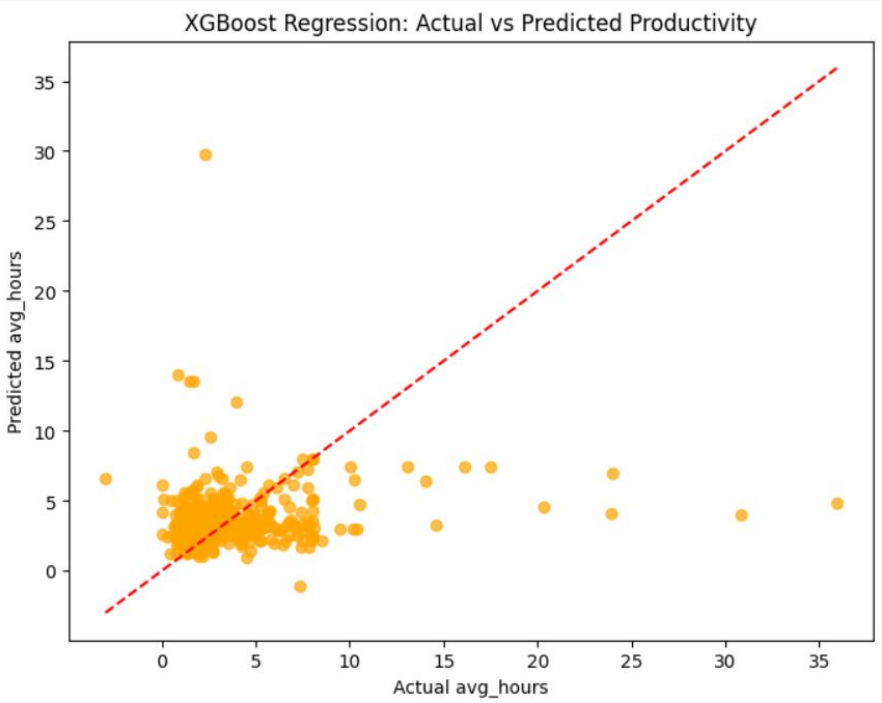


- 'page_rank'
- 'betweenness'
- 'distance_to_coach'

- 'time_entries'

Predicting Productivity





R^2 Score: -0.11128296311279051

R^2 Score: -0.16989419671483286

Insights

Retention

- Workload balance (avg_hours & time_entries) is the strongest predictor:
  - Employees with moderate, stable hours stay longer
  - Excessive or insufficient hours correlate with attrition
- Network centrality (PageRank, betweenness) has secondary influence
- Coaching/mentoring relationships show negligible impact

Productivity

- Insufficient data: Current features fail to model productivity effectively. More granular metrics needed

3. What factors contribute to project success or failure?

- 'num_employees'
- 'total_time_entries'

- 'billable_hours'
- 'total_hours'
- 'billable_ratio'

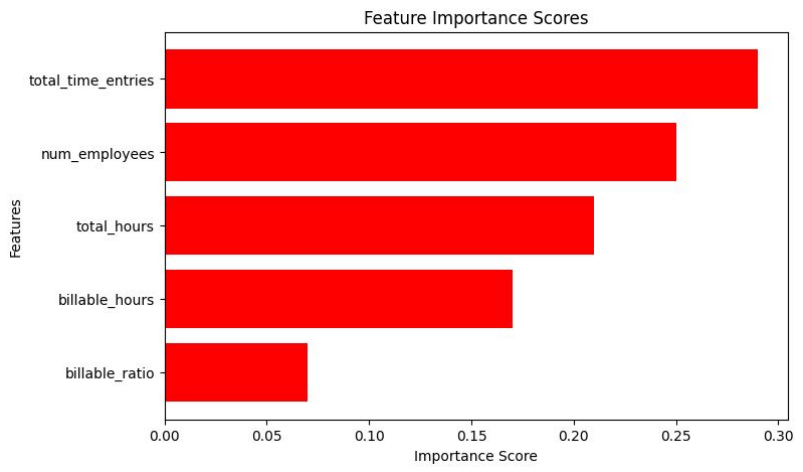
Random Forest vs. XGBoost

Random Forest Classifier Report:

	precision	recall	f1-score	support
0	0.71	0.73	0.72	1470
1	0.68	0.65	0.67	1279
accuracy			**0.69**	2749
macro avg	0.69	0.69	0.69	2749
weighted avg	0.69	0.69	0.69	2749

XGBoost Classifier Report:

	precision	recall	f1-score	support
0	0.71	0.74	0.72	1470
1	0.69	0.65	0.67	1279
accuracy			**0.70**	2749



Status

0: Active

1: Inactive or Dormant

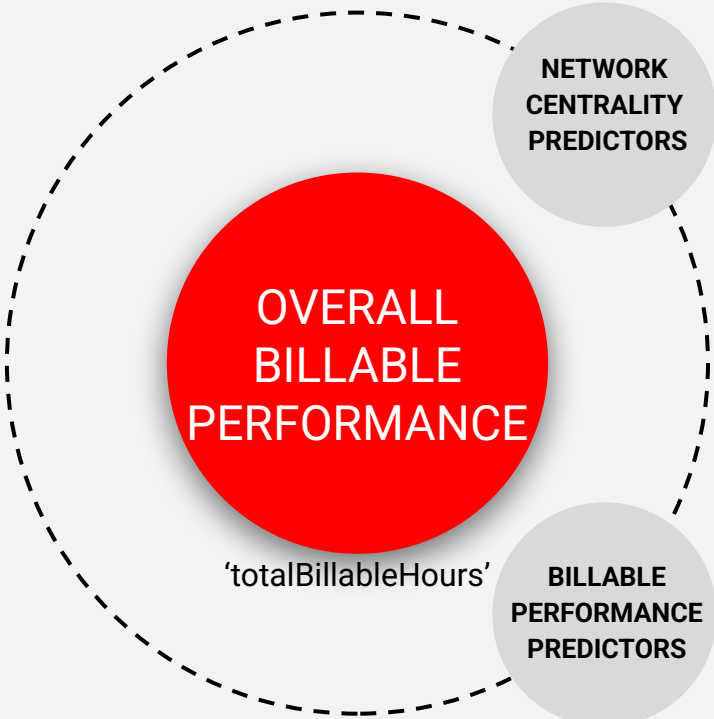
Insights

- **Total Time Entries (Most Important)** – Frequent work logging indicates engagement and accountability. Encourage consistent time tracking.
- **Number of Employees** – Larger teams contribute to success, but efficiency in task allocation is key.
- **Total Hours Worked** – More hours improve success, but productivity should be optimized to prevent burnout.
- **Billable Hours** – Revenue-generating work impacts success, but balance with non-billable tasks.
- **Billable Ratio (Least Important)** – Non-billable work (planning, collaboration) also plays a role in project outcomes.

Actionable Takeaways

- Promote consistent time tracking.
- Ensure adequate but efficient team sizes.
- Optimize work hours for efficiency, not just volume.
- Balance billable and non-billable work.

4. How does an employee’s centrality in the coaching network correlate with their overall billable performance?

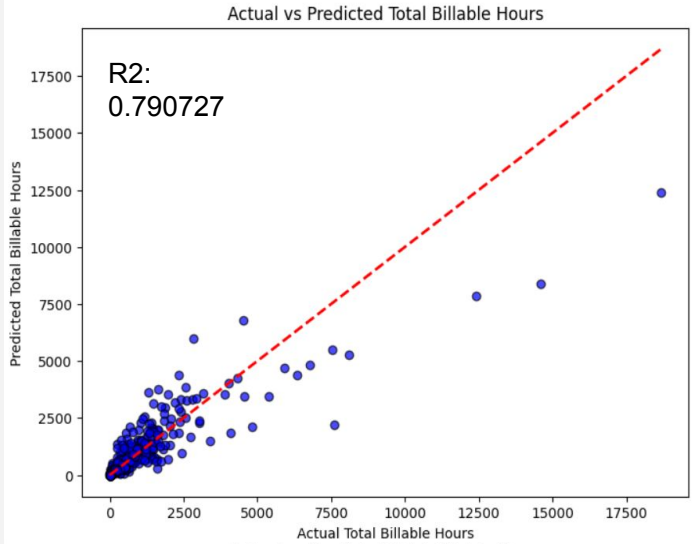


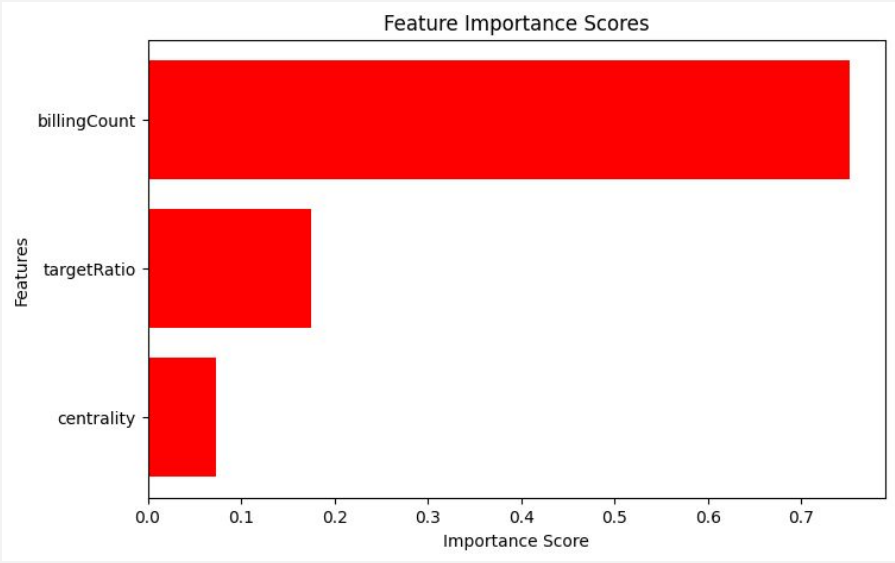
- 'centrality'

- 'billingCount'
- 'targetRatio'

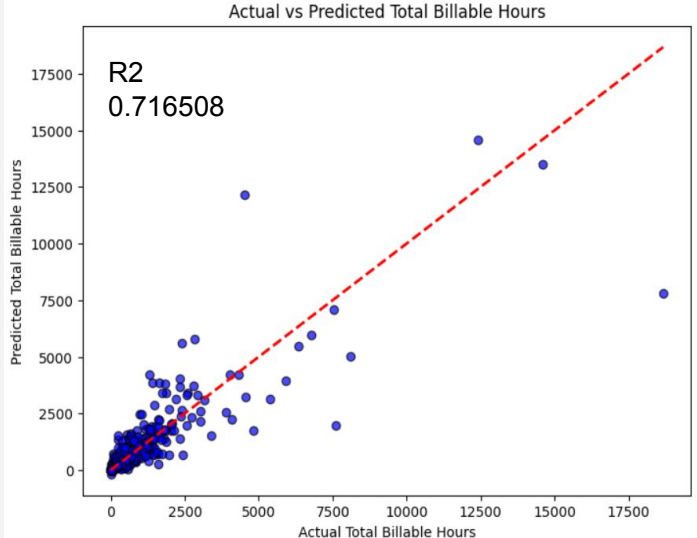
Random Forest vs. XGBoost

Random Forest





XGBoost



Insights

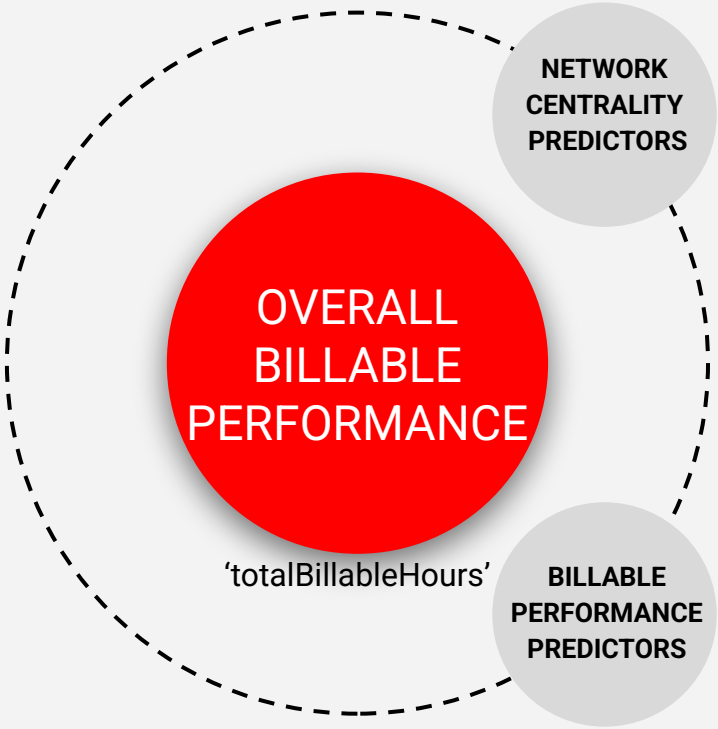
Key Findings

- **Billing activity dominance:** ‘billingCount’ is the primary driver of total billable hours, highlighting transaction volume as the strongest predictor.
- **Target expectations:** ‘targetRatio’ meaningfully contributes, reflecting alignment between organizational goals and employee outcomes.
- **Network centrality:** Coaching-network centrality shows statistical significance but minimal practical impact, suggesting secondary influence compared to direct billing metrics.

Implications

- Prioritize boosting productive billing activity and refining target-setting processes.
- Coaching networks may offer indirect benefits (e.g., long-term skill development) worth exploring through advanced modeling or longitudinal analysis.

5. How does partner status moderate the relationship between an employee's centrality in the coaching network and their overall billable performance?

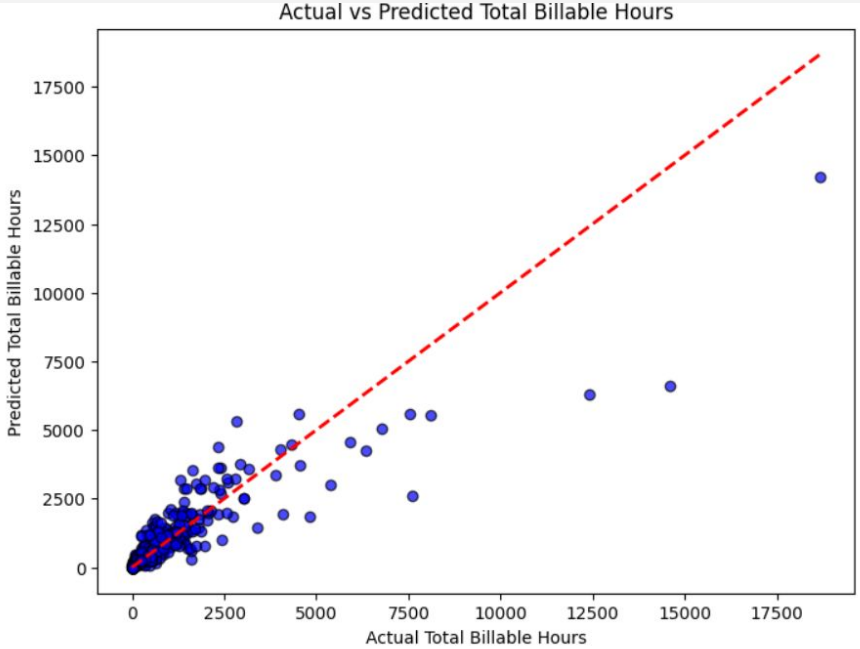


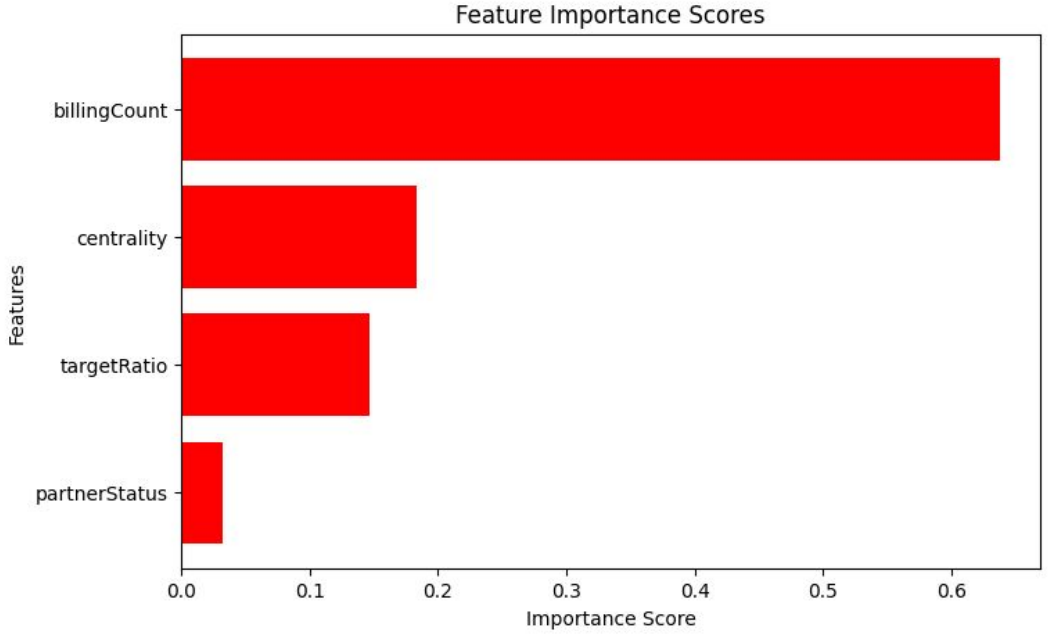
- 'centrality'

- 'billingCount'
- 'targetRatio'
- 'partnerStatus'

Random Forest Result

R2 : 0.7899489703164351





Insights

- Partner status may have strategic or cultural significance, but its impact on billable hours is minimal
- Billing transactions, network centrality, and target expectations are stronger predictors of billable hours
- Direct performance metrics outweigh partner status in influencing productivity

Conclusion

- **Employee attrition:** The strongest predictors of termination risk include billable rates, performance against targets, and network centrality. Employees with weak connections in the organization are at higher risk of termination.
- **Productivity insights:** Billing activity is the dominant driver of total billable hours, while target ratio alignment and network position play secondary roles.
- **Project success:** The number of employees, total time entries, and billable hours contribute significantly to project success, emphasizing the importance of workload balance and team efficiency.
- **Coaching network influence:** While coaching relationships impact employee engagement, their direct effect on productivity is minimal compared to billing-related metrics.
- **Partner status effect:** Being a partner does not significantly alter the predictive impact of network centrality on billable performance, indicating that other factors like billing activity hold greater weight.