

Validating logistic regression results

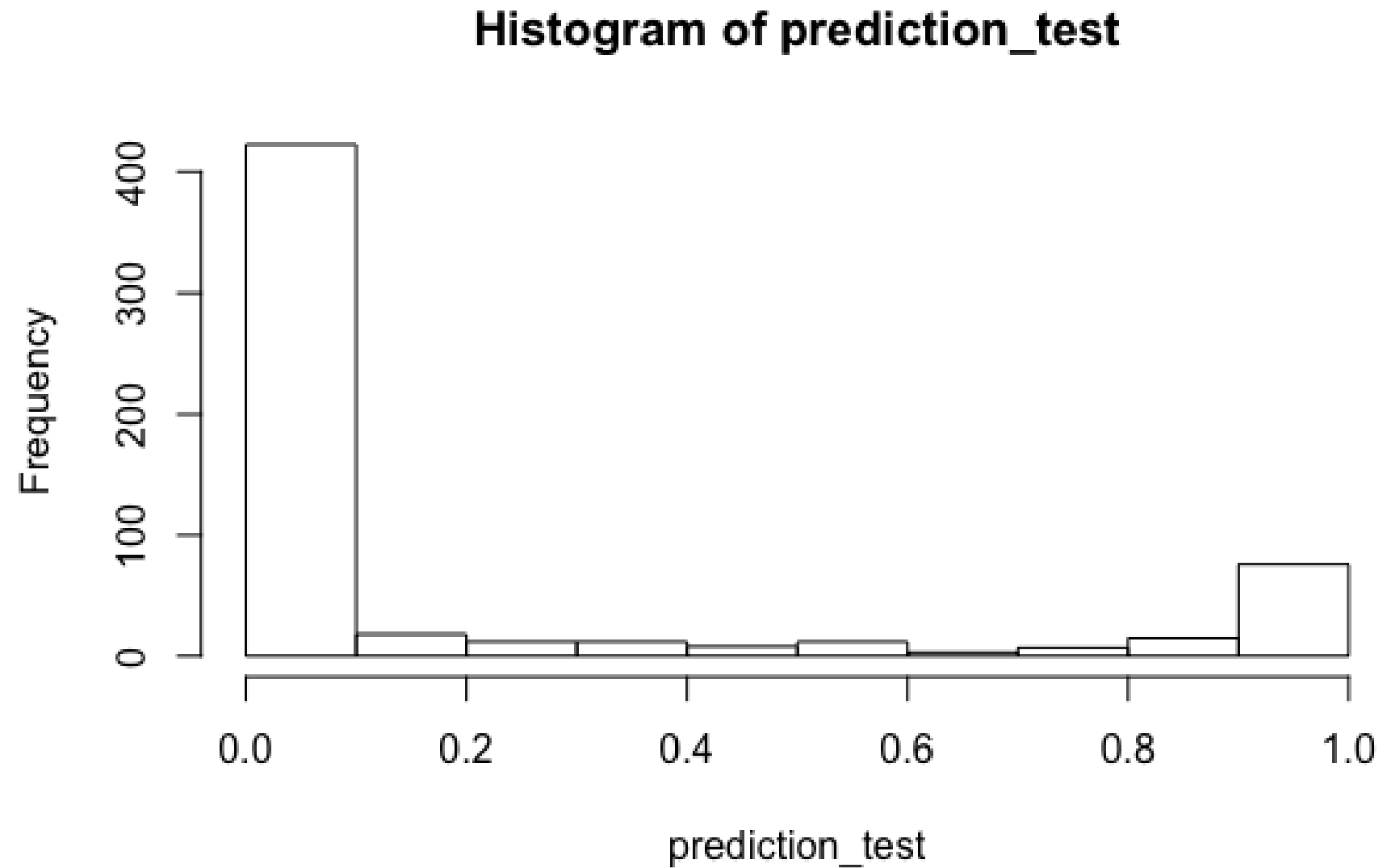
HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



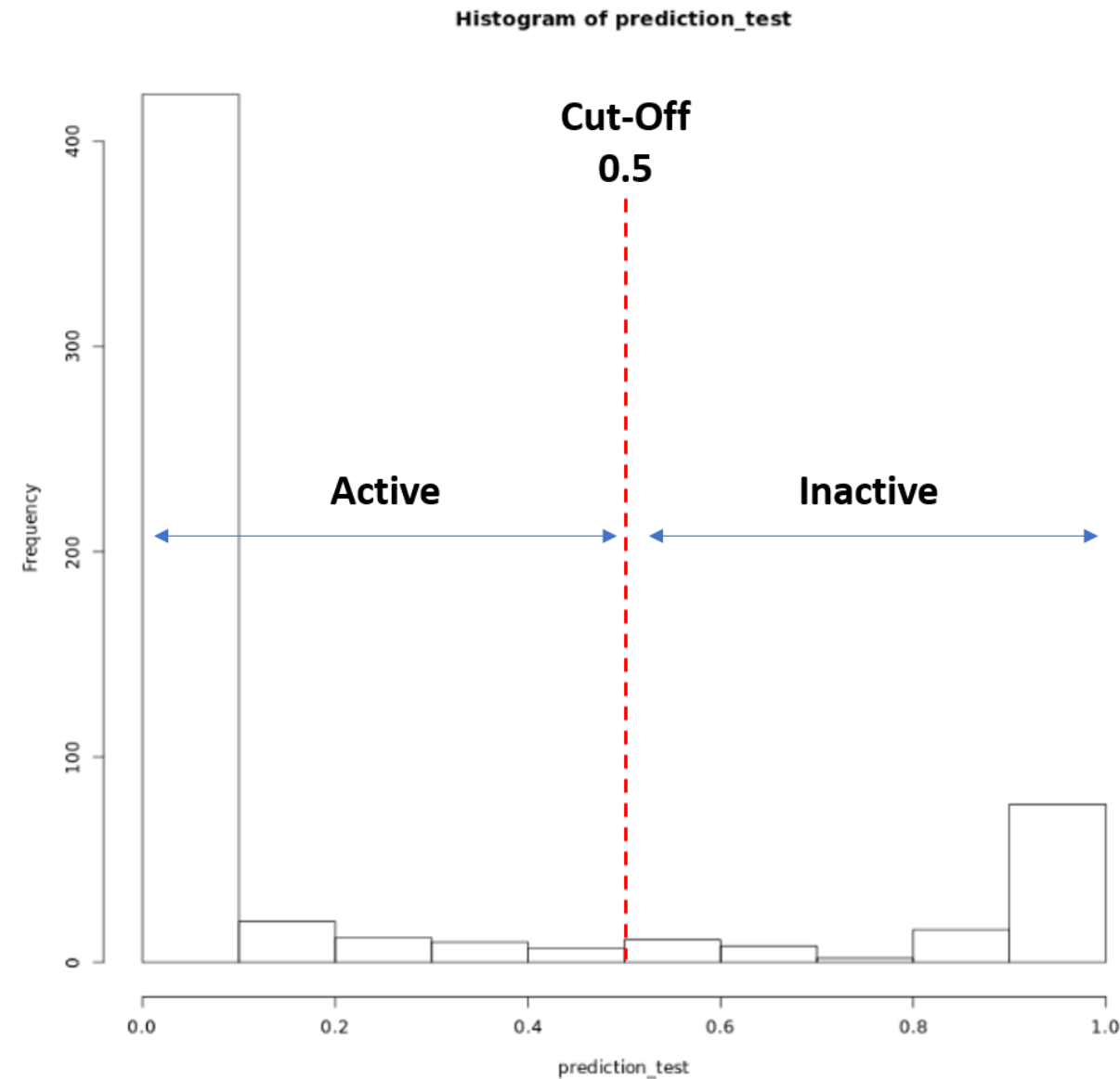
Anurag Gupta

People Analytics Practitioner

Turnover probability distribution of test cases



Turn probabilities in categories by using a cut-off



Turn probabilities in categories by using a cut-off

```
# Classify predictions using a cut-off of 0.5  
pred_cutoff_50_test <- ifelse(predictions_test > 0.5, 1, 0)
```

What is confusion matrix?

Confusion matrix measures the performance of a classification model.

Creating confusion matrix

```
## Creating confusion matrix  
table(pred_cutoff_50_test, test_set$turnover)
```

```
prediction_categories  0    1  
                    0 450  22  
                    1  20  94
```

Understanding confusion matrix

		Actual Employee Status		
		Actual: Active (0)	Actual: Inactive (1)	
Model Predicted	n = 586			
	Predicted: Active (0)	TN = 450	FN = 22	472
	Predicted: Inactive (1)	FP = 20	TP = 94	114
		470	116	
Cut-off: 0.5				

- True negatives (TN): The model correctly identified active employees
- True positives (TP): The model correctly identified inactive employees
- False positives (FP): The model predicted employees as inactive, but they are actually active
- False negatives (FN): The model predicted employees as active, but they are actually inactive

Confusion matrix: accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Accuracy} = \frac{450 + 94}{450 + 94 + 22 + 20}$$

=

0.9283

Creating confusion matrix

```
# Load library
library(caret)

# Construct a confusion matrix
conf_matrix_50 <- confusionMatrix(table(test_set$turnover,
                                         pred_cutoff_50_test))
```

```
conf_matrix_50
```

Confusion Matrix and Statistics

```
prediction_categories    0    1
                        0 450  22
                        1  20  94
```

Accuracy : 0.9283

95% CI : (0.9044, 0.9479)

No Information Rate : 0.802

P-Value [Acc > NIR] : <2e-16

Kappa : 0.7728

Mcnemar's Test P-Value : 0.8774

Sensitivity : 0.9574

Specificity : 0.8103

Pos Pred Value : 0.9534

Neg Pred Value : 0.8246

...

Resources for advanced methods

- [Supervised Learning in R: Classification](#)
- [Machine learning in the Tidyverse](#)

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R

Designing retention strategy

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta

People Analytics Practitioner

Know who may leave

```
# Load tidypredict
library(tidypredict)

# Calculate probability of turnover
emp_risk <- emp_final %>%
  filter(status == "Active") %>%
  # Add predictions using the final model
  tidypredict_to_column(final_log)
```

Know who may leave

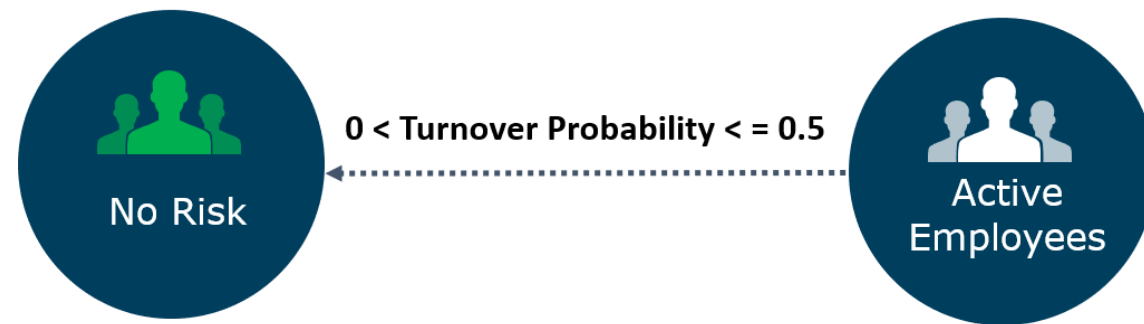
```
# Look at the employee's probability of turnover
emp_risk %>%
  select(emp_id, fit) %>%
  top_n(5, wt = fit)
```

```
# A tibble: 5 x 2
  emp_id    fit
  <chr>    <dbl>
E202     0.9694593
E6475     0.9814252
E6574     0.9983320
E7105     0.9193704
E9878     0.9371767
```

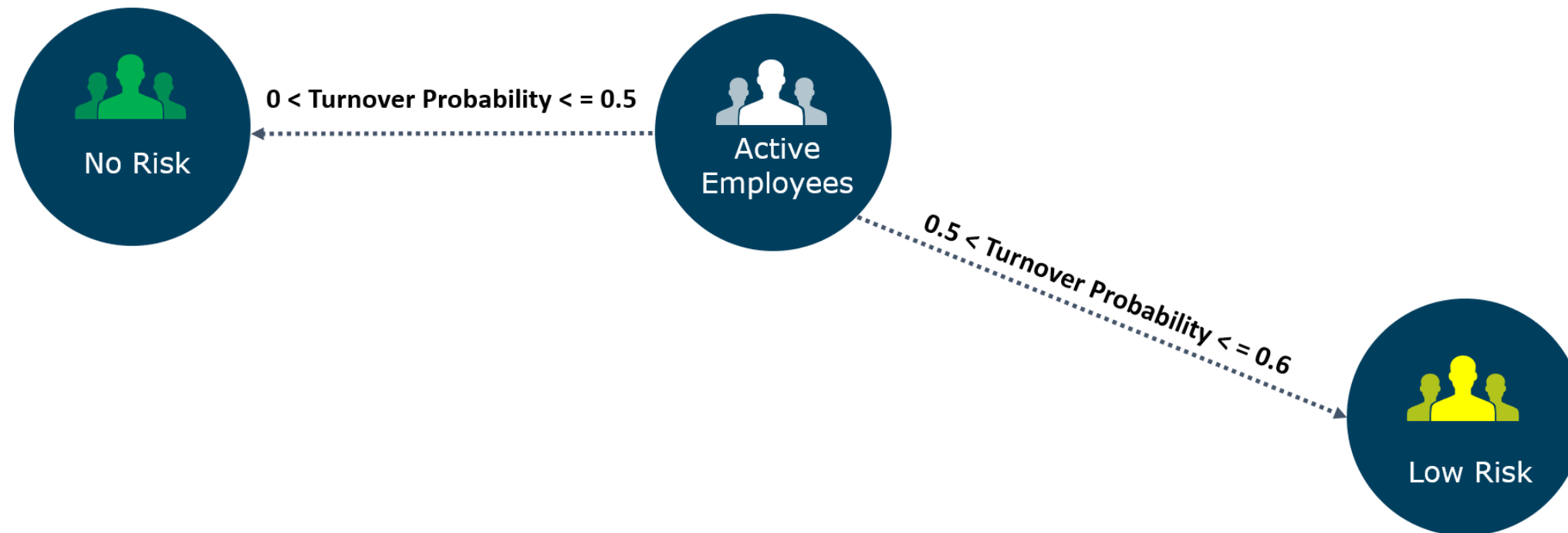
Classification of employees in risk buckets



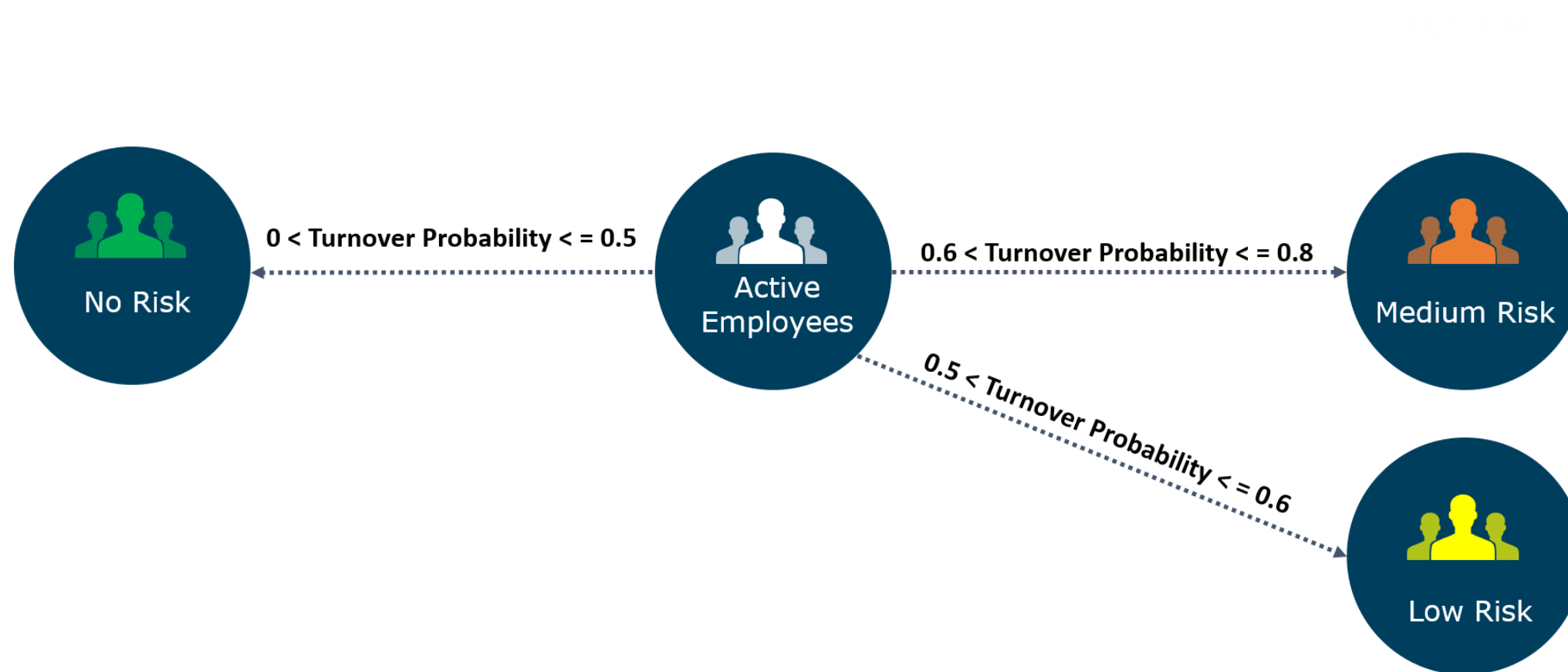
Classification of employees in risk buckets



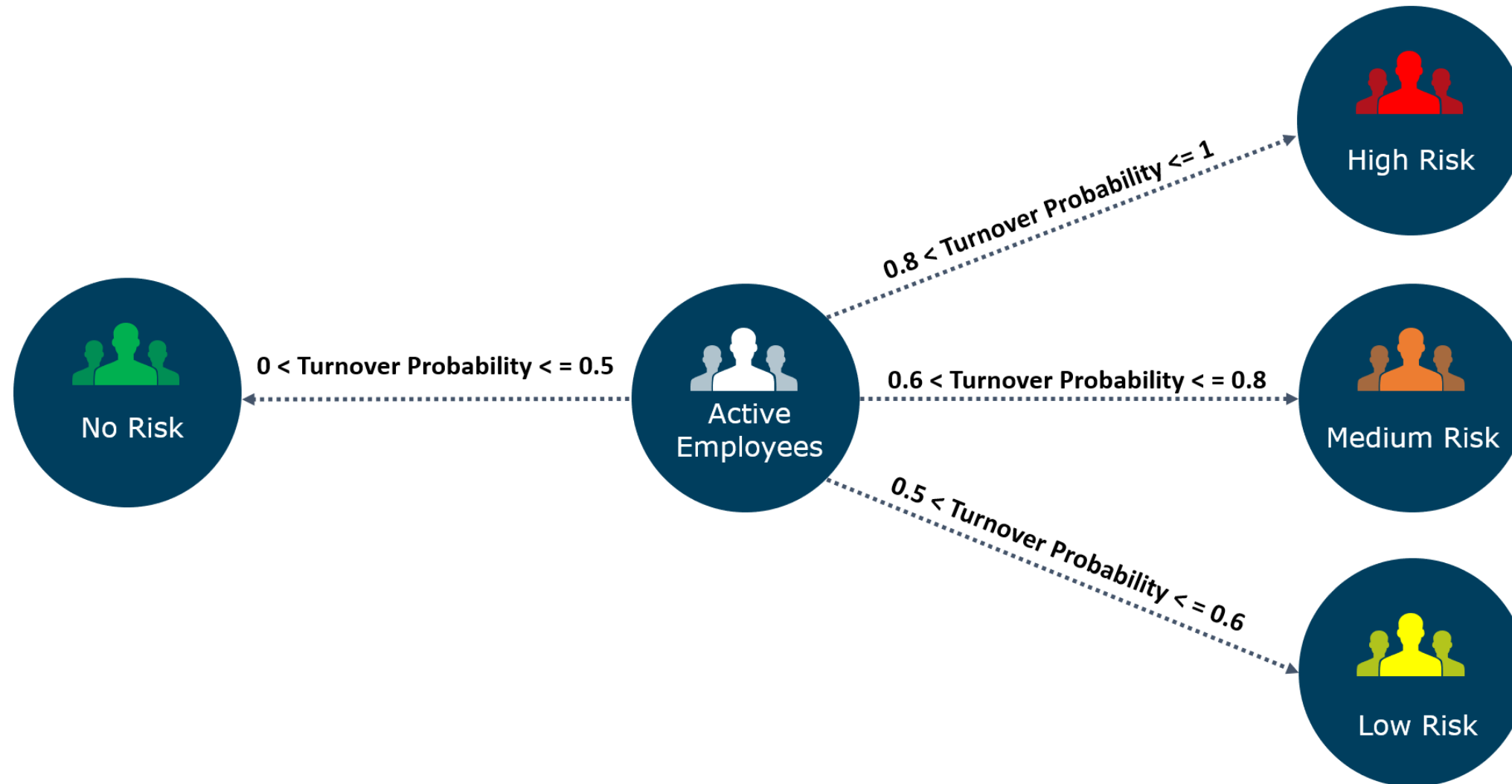
Classification of employees in risk buckets



Classification of employees in risk buckets



Classification of employees in risk buckets



Classify employees into risk buckets in R

```
# Create turnover risk buckets
emp_risk_bucket <- emp_risk %>%
  mutate(risk_bucket = cut(fit, breaks = c(0, 0.5, 0.6, 0.8, 1),
                           labels = c("no-risk", "low-risk",
                                       "medium-risk", "high-risk")))
```

Retention strategy

High Risk

- Immediate action planning
- Inform reporting manager
- Hold one-on-one conversation

Medium Risk

- Medium-term action planning
- Keep tracking for any behavioral change
- Have one-on-one or open house discussion

Retention strategy

Low Risk

- Long-term action planning
- Keep tracking for any behavioral change
- Have open house discussion

No risk

- No action required

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R

Return on investment calculation

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta

People Analytics Practitioner

Total cost of employee turnover

- Costs to off-board employee
- Cost-per-hire for replacement
- Transition costs, including opportunity costs

Understand the cost implication of high turnover rate

Turnover overview	Scenario 1	Scenario 2	% Change
Total Turnover	300	200	33%
Average Cost of Turnover**	\$40,000	\$40,000	0%
Total Cost of Turnover	\$12,000,000	\$8,000,000	\$4,000,000

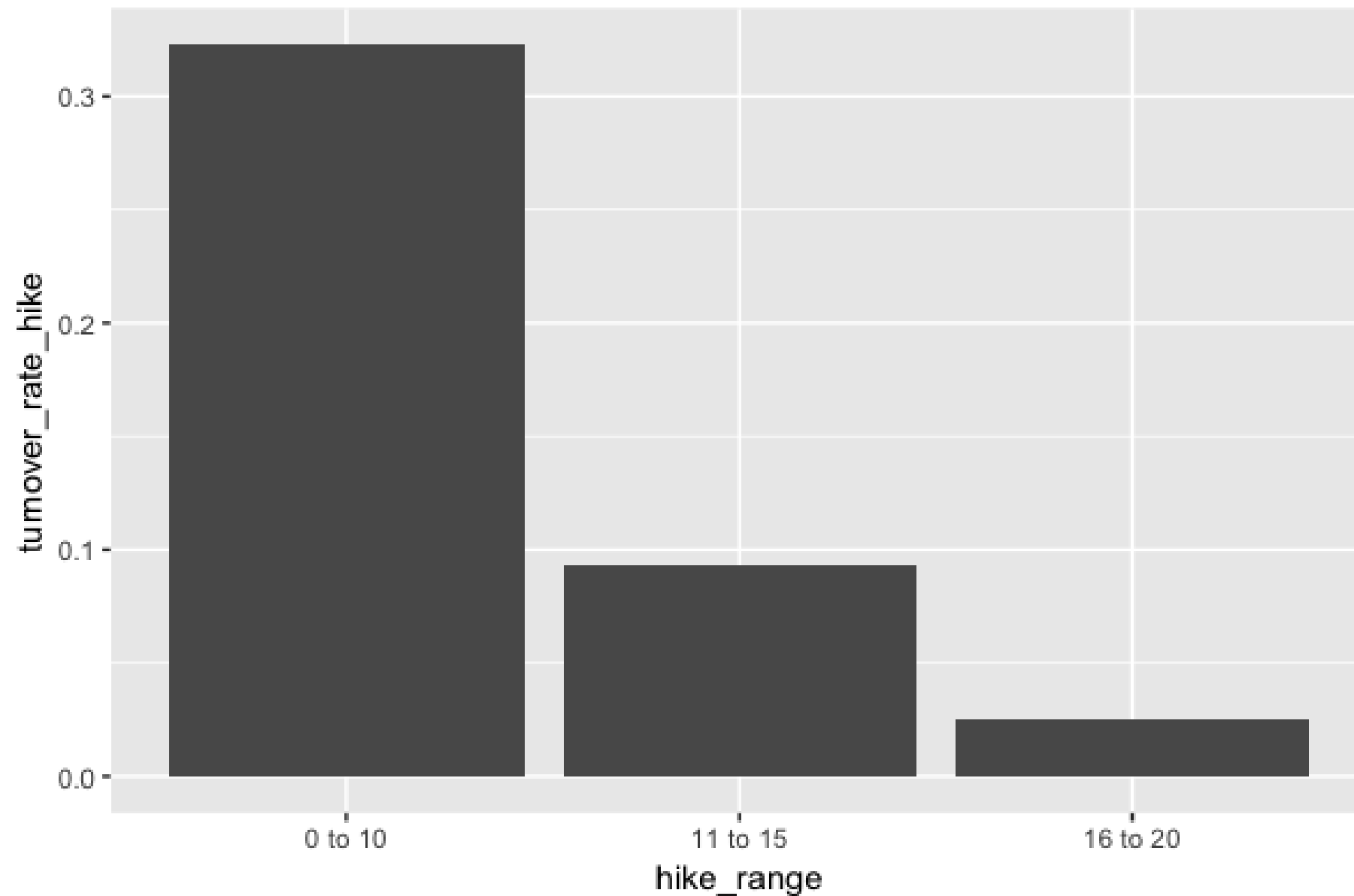
**[Enrich.org](https://enrich.org/)

Calculating ROI

$$\text{ROI} = \frac{\text{Program Benefits}}{\text{Program Cost}}$$

```
percent_hike          -0.59500    0.08134   -7.315  2.57e-13 ***
```

Turnover rate across salary hike range



Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R

Course Wrap-up

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R



Anurag Gupta

People Analytics Practitioner

Course Wrap-up

- What is employee turnover?
- HR data sources
- Derive new variables and variable importance
- Explore and validate
- Predict probability of turnover
- Designed retention strategies

**Go implement
employee turnover
prediction in your
organization!**

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN R