People Analytics

Project Report

at the Faculty of Business, Economics, and Law
Friedrich-Alexander-Universität Erlangen-Nürnberg
Schöller Endowed Chair for Information Systems
Prof. Dr. Sven Laumer

Presented by: Prabhu Nithin Gollapudi
Elisabethstraße 1
Erlangen
+49 176 60378999
prabhu.nithin.gollapudi@fau.de
23269243

Final Date of Submission: 2024-07-15

Content

1 Well-Being

1.1 The importance of well-being *(4 Points)*

The concept of well-being is seen as a broad and encompassing construct which has consequences at individual and organizational levels. The workforce is the company's most important asset and employee well-being has a direct link to performance in the workplace. The top three benefits of employers increasing their focus on employee well-being:

- A healthier and more inclusive culture.
- Better work-life balance.
- Better employee morale and engagement.

According to Pfeffer (2018), there are seven factors that directly impact the health of employees on the job:

1. Job design, including control over work - The negative affects of work are particularly acute for employees in high-pressure jobs with little control over their workdays.
2. Overtime & number of hours worked - Long work hours are associated with adverse health, including cardiovascular disease, diabetes, disability.
3. Providing social support - Relationships have a direct effect on health and buffer the effects of various psycho social stresses, including workplace stress, that compromises health.
4. Conflicts between family & work commitments - If employees are happy at home, they will bring their best selves to work every day.
5. Perceived fairness & justice at work – Not treating employees fairly has a negative impact on employee morale.
6. Layoffs & economic insecurity – These have become a very relevant factor because of mass layoffs, employees loose their confidence in the company.
7. Health insurance plans – A sense of medical security has a positive influence on employees.

FAU Bank can introduce the following organizational policies to monitor and bring positive influence in employee well-being:

- Flexible working hours
- Health and wellness programs
- Childcare benefits or services
- Taking leaves
- Organizational understanding and support
- Availability and usage of WLB policies

These policies help FAU Bank impact Employees' ability to achieve WLB as it plays the central role in attaining workplace health and well-being.

1.2 Stress induced by technology *(7 Points)*

Since FAU Bank is introducing a new piece of technology, this results in Technostress. Technostress is defined as stress that individuals experience due to the use of information and communication technologies (ICT). It happens when a person is exposed to a challenging or overwhelming situation in relation to ICT.

This introduction can lead to Challenge stressors and Hindrance stressors. We can use The Challenge-Hindrance Framework (Tarafdar 2019) to better understand the consequences.

- In challenge stressors, Demands are considered as obstacles to be overcome. Where as in hindrance stressors, Demands are considered stressful and prevent one's growth.
- Challenge stressors can be motivating, because people see room to grow. Hindrance stressors, can be non motivating, because people believe that the time and effort, they must put in, is not in equilibrium with the demands.
- Challenge stressors are positively related to performance. Hindrance stressors are negatively related to performance.

The bank tellers can view the change of replacing their legacy Transaction Processing System as a challenge stressor or as a hindrance stressor (Bala & Venkatesh 2015) which FAU Bank should focus on finding because

- If the user deems it to be a challenge stressor, the user sees the demands as an opportunity for himself to grow.
- If it is a hindrance stressor the demands bring a possibility for loss, restriction, or damage.

Following the Challenge and Threat Coping Behaviors (Beaudry & Pinsonneault 2005), FAU Bank should increase awareness of it and access via Appraisal whether it is an Opportunity or a Threat of High Control or Low control. Follow different adaptation strategies based on the previous analysis.

- Benefits Maximizing which results in individual efficiency and effectiveness in case of high control Opportunity.
- Benefits Satisficing which results in Individual efficiency and effectiveness in case of low control Opportunity.
- Disturbance Handling which results in Individual efficiency and effectiveness, Minimization of the negative consequences of the IT Event, and Restoring personal emotional stability in case of high control Threat.
- Self Preservation leading to Minimization of the negative consequences of the IT Event, Restoring personal emotional stability and even in some cases exiting the adaptation entirely in case of low control Threats.

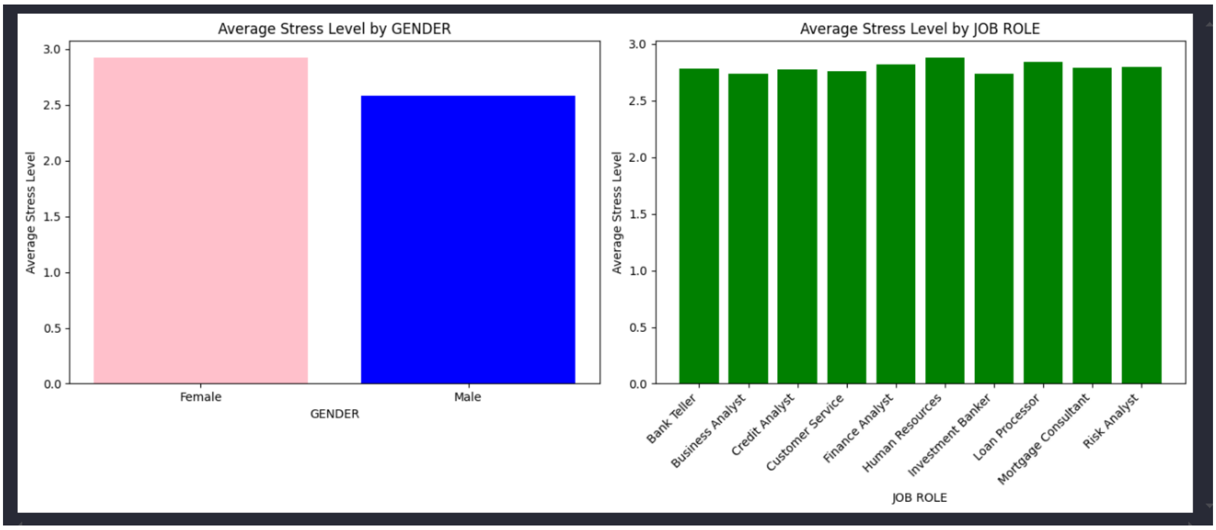
In any case FAU bank should ensure that Technostress leads to positive consequences of bank tellers.

1.3 Employee well-being dataset *(8 Points)*

The provided data on employee well-being is a simple structured and tabular format in csv format. The following preprocessing steps are performed on it:

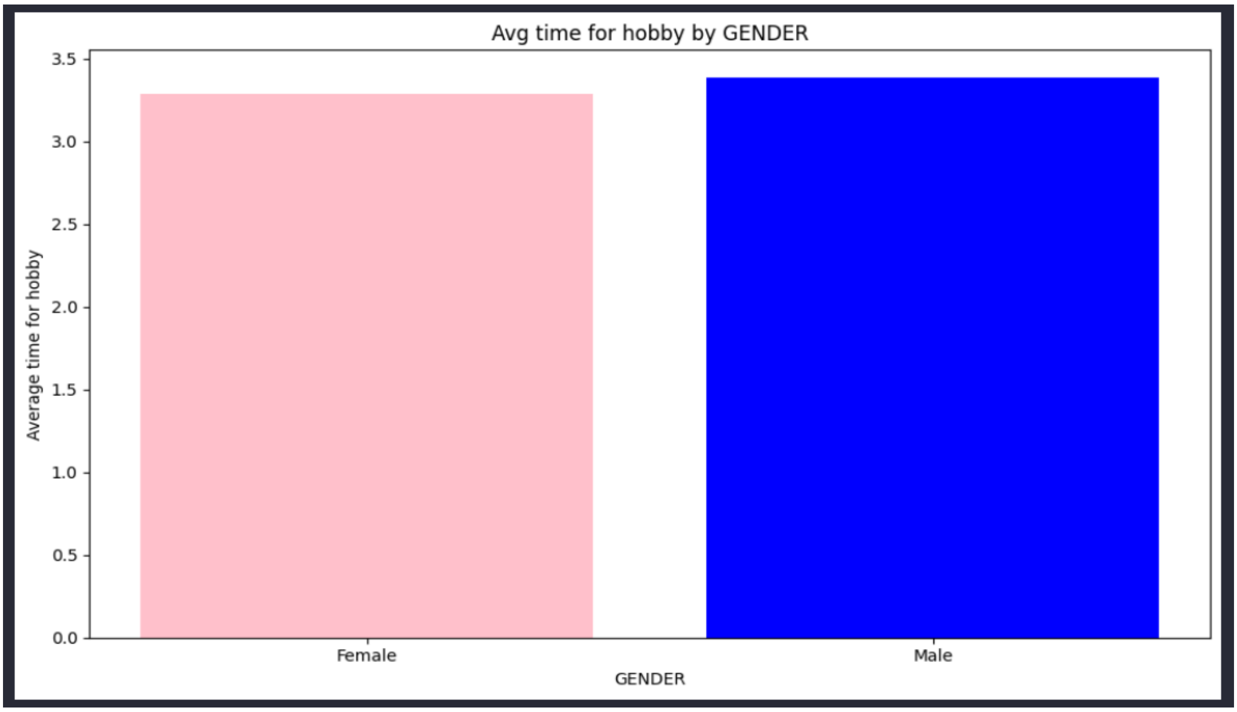
- Checking for missing values using - df.isnull().values.any() which returned True
- Dropping those values with missing entries. - df.isnull().values.any() returned False
- Dropped Employee_ID column from the analysis part because it does not have any impact on employee well-being.
- Data transformation on following columns – AGE, JOB_ROLE, GENDER. Converted these categorical data to numeric mapping given below.
  {'AGE': {0: '21 to 35', 1: '36 to 50', 2: '51 or more', 3: 'Less than 20'}, 'JOB_ROLE': {0: 'Bank Teller', 1: 'Business Analyst', 2: 'Credit Analyst', 3: 'Customer Service', 4: 'Finance Analyst', 5: 'Human Resources', 6: 'Investment Banker', 7: 'Loan Processor', 8: 'Mortgage Consultant', 9: 'Risk Analyst'}, 'GENDER': {0: 'Female', 1: 'Male'}}

The bar charts showing daily stress according to gender and job role is shown below.



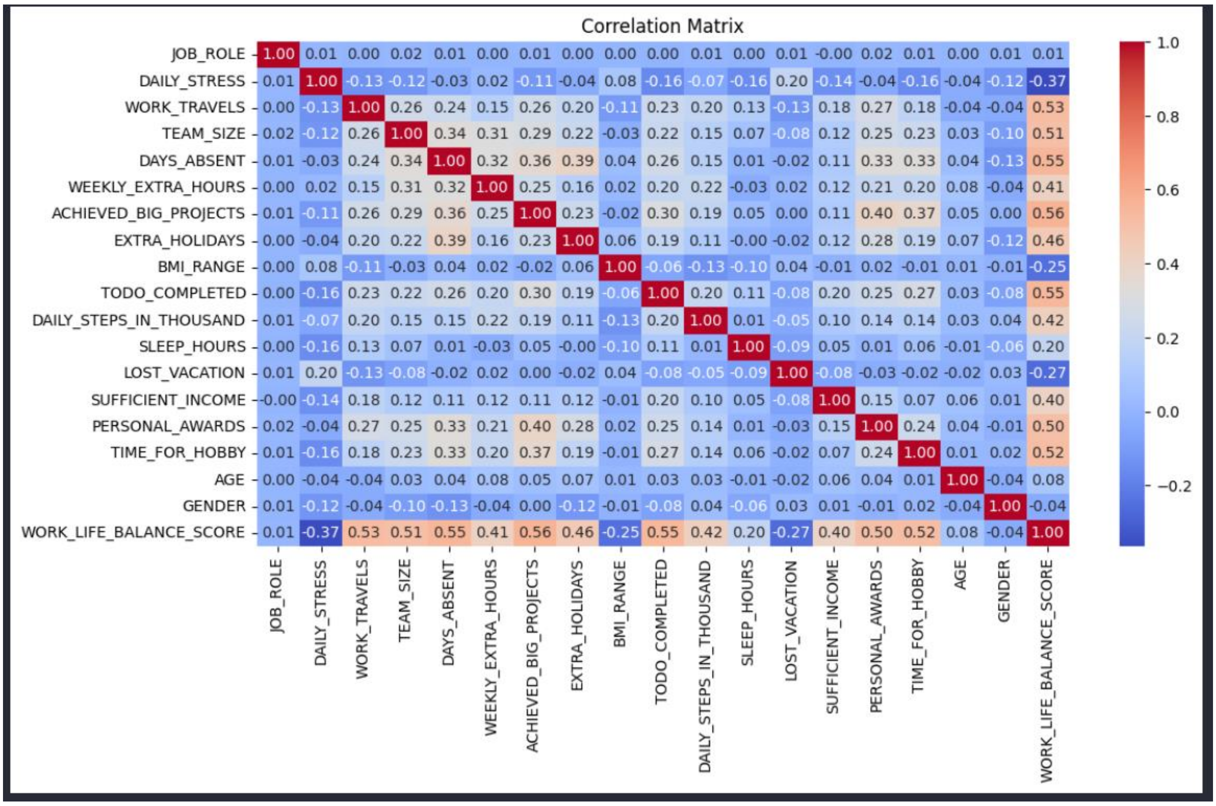
From the bar chart it is evident that on an average, female employees experience more stress than males and coming to job role, Human Resources experience more stress than other roles.

The average time spent for hobbies is more in males than in female as seen in the bar chart. Males spend around 3.3845 hours per week to their preferred hobby whereas females spend around 3.29 hours per week.



The correlation matrix helps us understand which factors correspond to high correlation with work life balance score.

The factors – WORK_TRAVELS, TEAM_SIZE, DAYS_ABSENT, ACHIEVED_BIG_PROJECTS, TODO_COMPLETED, PERSONAL_AWARDS AND TIME_FOR_HOBBY have a very high correlation to WORK_LIFE_BALANCE_SCORE



1.4 Predictive well-being algorithm *(6 Points)*

A Linear Regression model is trained on the given data set by considering all the features in the data set except the Employee_ID column.

1. The r2_score is 0.9449331698659169 which indicates that our model is doing a very good job of predicting the work life balance score of employees.

```
from sklearn.metrics import r2_score
r2_score(y_test, y_pred)
```

✓ 0.0s

```
0.9449331698659169
```

2. The difference between the real value and predicted value for 5 entries is as follows.

```
# a tabular form the difference between the real value found in the dataset and th

df_results = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df_results['Difference'] = df_results['Actual'] - df_results['Predicted']
df_results.head()
```

✓ 0.0s

	Actual	Predicted	Difference
5476	632.6	627.244123	5.355877
4507	652.0	659.840253	-7.840253
8812	647.5	655.455634	-7.955634
11151	622.3	639.515359	-17.215359
5242	780.4	773.786772	6.613228

The difference between actual value and predicted value is not big in comparison to the actual values (deviation is approximately in range 1 to 3 %).

3. Prediction of wlb score of new employee with features 33 year old male working as a bank teller came around 630.39.
Forming the input feature vector –
JOB_ROLE – 0, DAILY_STRESS – 2, WORK_TRAVELS – 3, TEAM_SIZE – 4, DAYS_ABSENT – 5, WEEKLY_EXTRA_HOURS – 7, ACHIEVED_BIG_PROJECTS – 2, EXTRA_HOLIDAYS – 2, BMI_RANGE – 2, TODO_COMPLETED – 3, DAILY_STEPS_IN_THOUSAND – 5, SLEEP_HOURS – 7, LOST_VACATION – 5, SUFFICIENT_INCOME – 2, PERSONAL_AWARDS – 4, TIME_FOR_HOBBY – 2, AGE – 0, GENDER – 1.

Predict the WLB score for a new employee

```python
# predicting the work life balance score for a new employee

predicted_value = ml.predict([[0, 2, 3, 4, 5, 7, 2, 2, 2, 3, 5, 7, 5, 2, 4, 2, 0, 1]])
print(predicted_value)
```

0.0s

```
[630.3839036]
c:\Users\prabh\FAU\Study\SoSe2024\PA\.venv\Lib\site-packages\sklearn\base.py:493: UserWarning: 
  warnings.warn(
```

Based on the regression model, it can be said that the features – DAILY_STRESS, BMI_RANGE, SLEEP_HOURS, SUFFICIENT_INCOME , GENDER have very high positive coefficients whereas LOST_VACATION has a negative coefficient.

```python
import statsmodels.api as sm
model = sm.OLS(y, x).fit()
print(model.summary2())
```

0.2s

Results: Ordinary least squares

Model:	OLS	Adj. R-squared (uncentered):	0.992
Dependent Variable:	WORK_LIFE_BALANCE_SCORE	AIC:	176076.2669
Date:	2024-07-11 14:06	BIC:	176214.4804
No. Observations:	15971	Log-Likelihood:	-88020.
Df Model:	18	F-statistic:	1.095e+05
Df Residuals:	15953	Prob (F-statistic):	0.00
R-squared (uncentered):	0.992	Scale:	3589.4

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
JOB_ROLE	3.1056	0.1626	19.1053	0.0000	2.7869	3.4242
DAILY_STRESS	12.1914	0.3461	35.2286	0.0000	11.5131	12.8697
WORK_TRAVELS	2.4337	0.1598	15.2310	0.0000	2.1205	2.7469
TEAM_SIZE	3.9701	0.1890	21.0031	0.0000	3.5996	4.3406
DAYS_ABSENT	2.4997	0.1793	13.9403	0.0000	2.1482	2.8512
WEEKLY_EXTRA_HOURS	3.0640	0.1712	17.9003	0.0000	2.7285	3.3995
ACHIEVED_BIG_PROJECTS	1.7269	0.2068	8.3527	0.0000	1.3217	2.1322
EXTRA_HOLIDAYS	4.8807	0.2874	16.9835	0.0000	4.3174	5.4440
BMI_RANGE	27.0250	0.9208	29.3507	0.0000	25.2202	28.8298
TODO_COMPLETED	4.6143	0.2030	22.7317	0.0000	4.2164	5.0122
DAILY_STEPS_IN_THOUSAND	5.8388	0.1723	33.8927	0.0000	5.5011	6.1765
SLEEP_HOURS	37.8929	0.2935	129.1280	0.0000	37.3177	38.4681
LOST_VACATION	-0.0072	0.1319	-0.0546	0.9565	-0.2657	0.2513
SUFFICIENT_INCOME	68.0043	1.0419	65.2695	0.0000	65.9620	70.0465
PERSONAL_AWARDS	2.5743	0.1770	14.5397	0.0000	2.2272	2.9213
TIME_FOR_HOBBY	3.1861	0.1967	16.1969	0.0000	2.8005	3.5717
AGE	6.1950	0.4662	13.2880	0.0000	5.2812	7.1088
GENDER	25.0376	0.9918	25.2450	0.0000	23.0936	26.9816

FAU Bank should consider focusing on taking care of employee personal well-being by providing more vacations (making up for lost_vacations), conducting physical activities which benefit employee health (bmi_range, sleep_hours) and finally making sure that their current compensation is being maintained (SUFFICIENT_INCOME). Special focus needs to be considered while implementing this as GENDER affects the parameters.

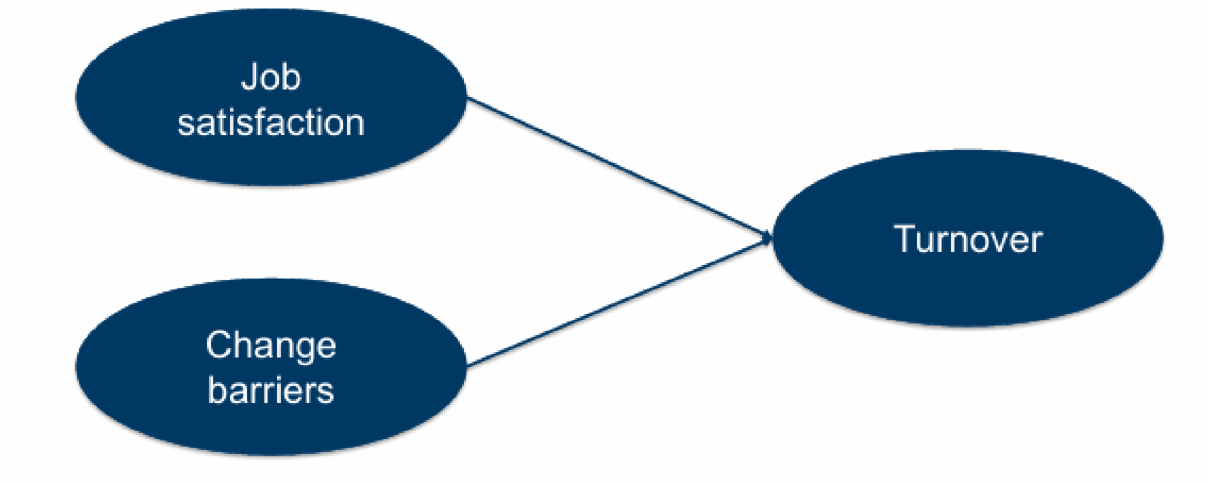
2 Turnover

2.1 Employee Turnover *(4 Points)*

Voluntary turnover or resignation reflects an employee's decision to leave an organization. If FAU Bank is reaching concerning levels of turnover rate, then it results in the following:

- It is extremely costly for the company, as it usually involves the loss of a high-performing employee.
- In addition to the loss of knowledge, replacing an employee often costs between 30% and 400% of the respective annual salary.

According to The Organization Equilibrium Theory (March and Simon 1958), the underlying assumption is that turnover occurs when individuals perceive that their own contributions within an organization exceed the incentives they receive from that organization. It also states that Job satisfaction and change barriers are leading causes of employee turnover.



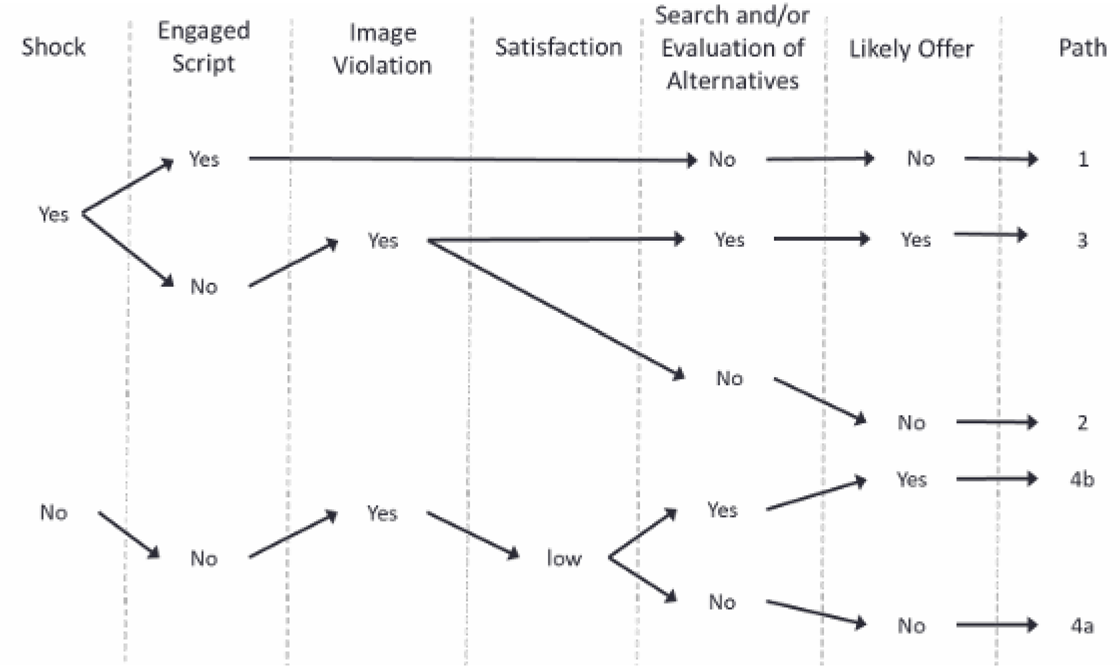
2.2 Employee Turnover Theories *(6 Points)*

The behavior of the bank teller, Jane, can be expressed by the Job Embeddedness Theory.

Job Embeddedness Theory explains why individuals are reluctant to leave organizations. The theory shows that individuals may remain in organizations because they are strongly socialized and reluctant to give up this position. It is at various dimensions Community fit, Fit to the organization, Community connection, Connection to the organization, Community related waiver and Organization-related waiver.

As Jane has created limited social connections and meaningful relationships in the workplace, she feels disconnected which can be seen affected in the dimensions of Community fit, Community connection. As she does not have any strongly socialized and reluctance to stay in the bank teller position, she is more likely to quit.

The "Unfolding Model of Turnover" (Lee and Mitchell 1994) can be used to explain why John wants to leave the company. According to it, there are four psychological paths when people leave a company.



As the company that John was working for was acquired by FAU Bank, this caused a shock to John as it triggered his independence and autonomy values. It can be understood that John may have followed Path 1, as he was directly impacted by the decision which has caused him shock.

2.3 Employee Turnover Dataset *(10 Points)*

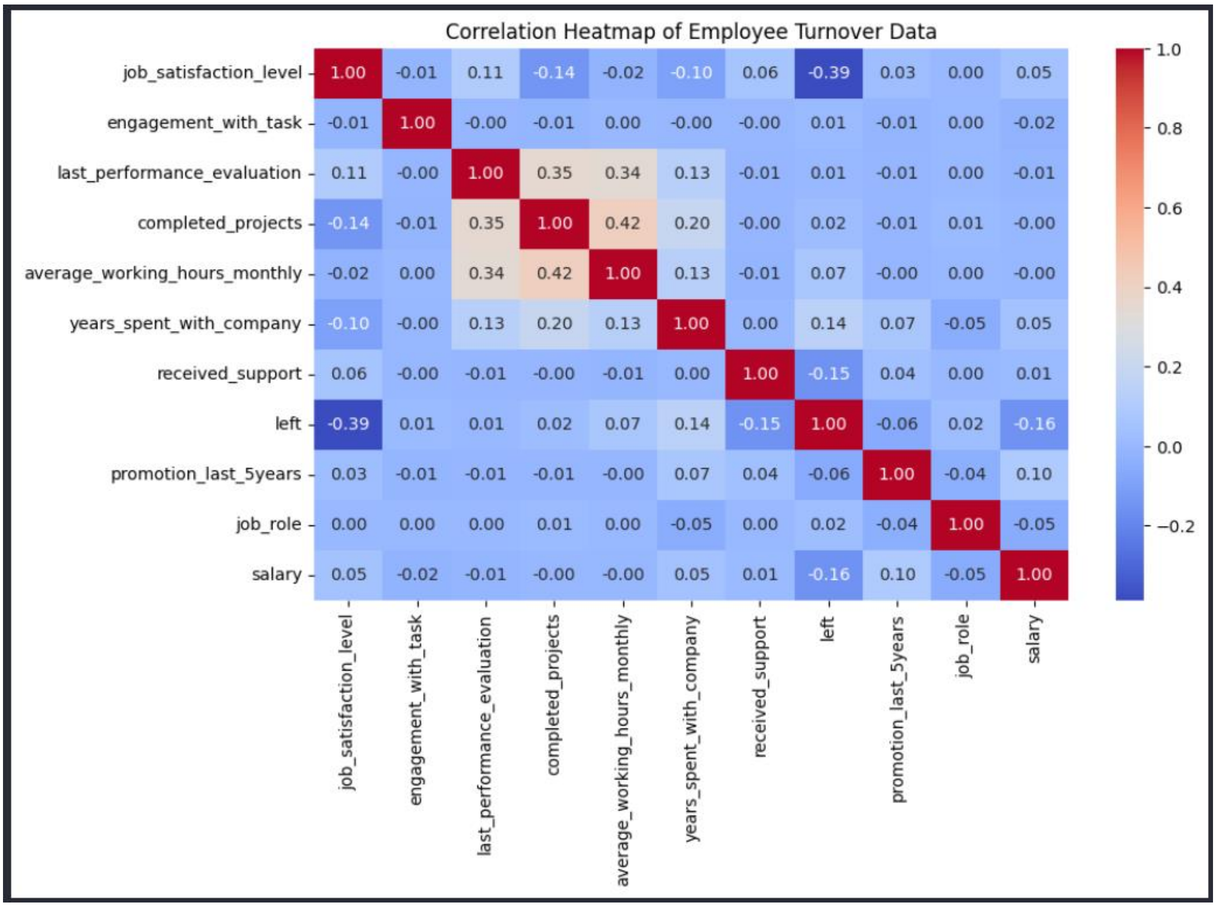
- The average job satisfaction level of employees who left FAU Bank is **0.44**. This is obtained by filtering all the employees who left using the left flag
  - **employees_left = df[df["left"] == 1]**
- The average salary satisfaction level of employees who left FAU Bank is **0.415.** We can interpert this result as – avg salary level of employees who left is in between (**low and medium**). This is obtained by filtering all the employees who left the organisation and label encoding the salary column using mapping.
  - **employees_left = df[df["left"] == 1]**
  - **salary_levels = {'low': 0, 'medium': 1, 'high': 2}**
- The employees who have left FAU Bank have stayed with the company for an average of **3.87** years. This is obtained by taking the average of all employees who left the company.

```
Average job satisfaction level for employees who left:  0.4400980117614114
Salary levels:  {'low': 0, 'medium': 1, 'high': 2}
Average salary level for employees who left:  0.41472976757210867
Average years spent with company for employees who left:  3.876505180621675
```

- Let us compare how the salaries are distributed between those who left the company and those who stayed with FAU Bank to see if there is any impact. From the graph it can be seen that the percentage of employees who leave the company is very high in low income levels than in high income levels.
    - In Low income levels, 29.69% of employees are likely to leave.
    - In Medium income levels, 20.43% of employees are more likely to leave and
    - In High income levels, only 6.63% are likely to leave FAU Bank.

So, we can say that salary level does affect if employees want to leave FAU Bank.

- To understand which attributes have a high correlation with the column ‘left‘, we can draw a heatmap and visualise it. To do this, we have to first convert the job_role column to numerical data. If FAU Bank might want to focus on improving job satisfaction level, salary and make sure employees receive more support.
  - The following attributes have a good positive correlation with left column – years_with_company – 0.14, average_working_hours_monthly - 0.07.
  - The following attributes have a very good negative correlation with left column – job_satisfaction_level -0.39 , received_support -0.15, salary -0.16.



Data preprocessing:

- Job roles is converted to numeric data using the label encoding technique and it resulted in the following map
  - Job Roles Encoding: {'IT': 0, 'bank_teller': 1, 'business_analyst': 2, 'credit_analyst': 3, 'customer_service': 4, 'finance_analystD': 5, 'hr': 6, 'investment_banker': 7, 'loan_analyst': 8, 'mortgag_consultant': 9}
- Salary is converted to numeric data using a constant mapping
  - salary levels: {'low': 0, 'medium': 1, 'high': 2}
- Transforming continuous data by applying data binning
  - Job satisfaction is transformed into five different bins with values - [0, 1, 2, 3, 4]
  - Last performance evaluation is changed to 5 different bins with values - [0, 1, 2, 3, 4].
- The columns number of projects and working hours is merged into one new feature –
  - Created new column - normalized_score by normalizing and adding both.
  - df['normalized_score'] = scaler.fit_transform(df[['completed_projects']]) + scaler.fit_transform(df[['average_working_hours_monthly']]).

2.4 **Turnover prediction** ***(5 Points)***

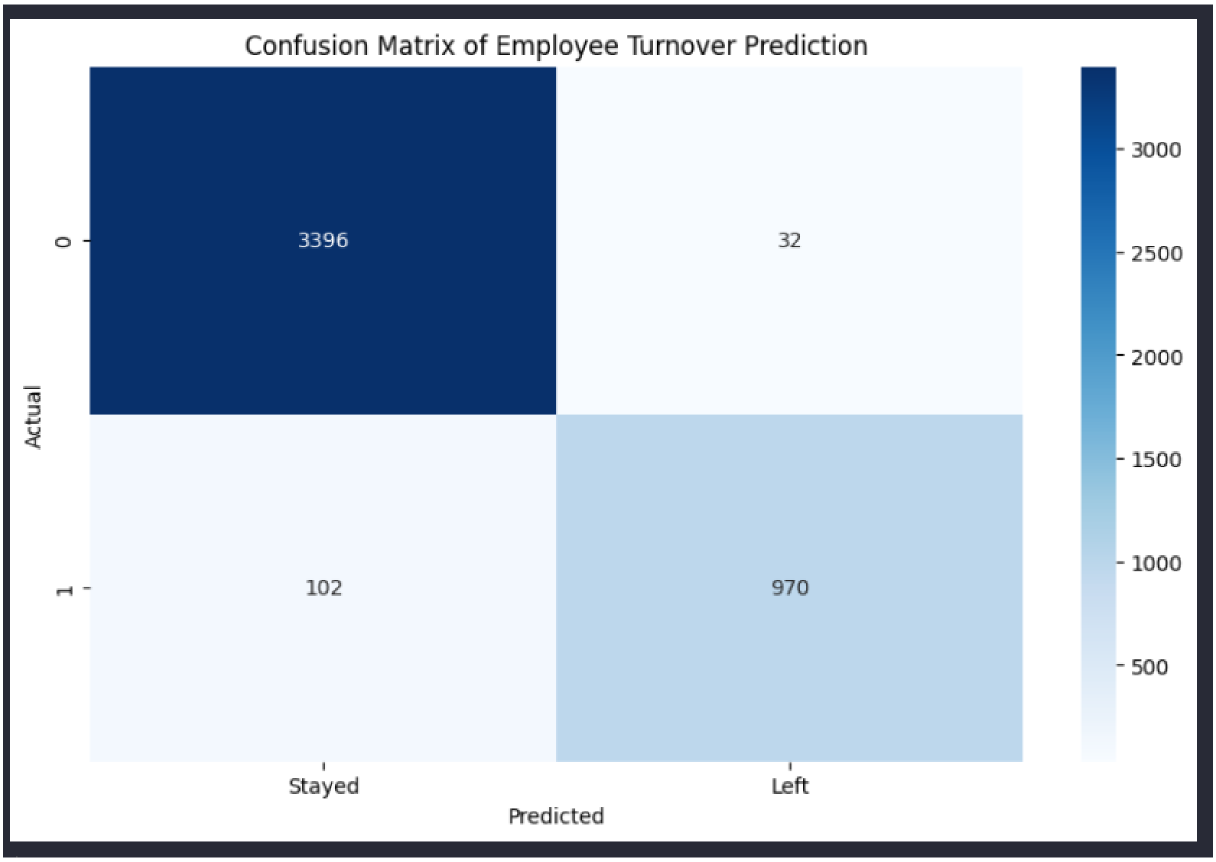
The dataset is trained on a Random Forest Classification algorithm which is an ensemble technique where the best of classifications is selected.

- The input feature vector is
    - job_satisfaction_level, engagement_with_task, last_performance_evaluation, completed_projects, average_working_hours_monthly, years_spent_with_company, received_support, promotion_last_5years, job_role, salary, normalized_score.
- The output variable is
    - Left

The choice of classification algorithm is random forest because of the following reasons

- Ensemble learning – It combines multiple results and does a better job at generalizing to new data.
- Robust to noise – It is less sensitive to noise as the average effect takes care of it.
- Reduce Overfitting – random forest reduces the risk of overfitting and stabilizes the predictions.

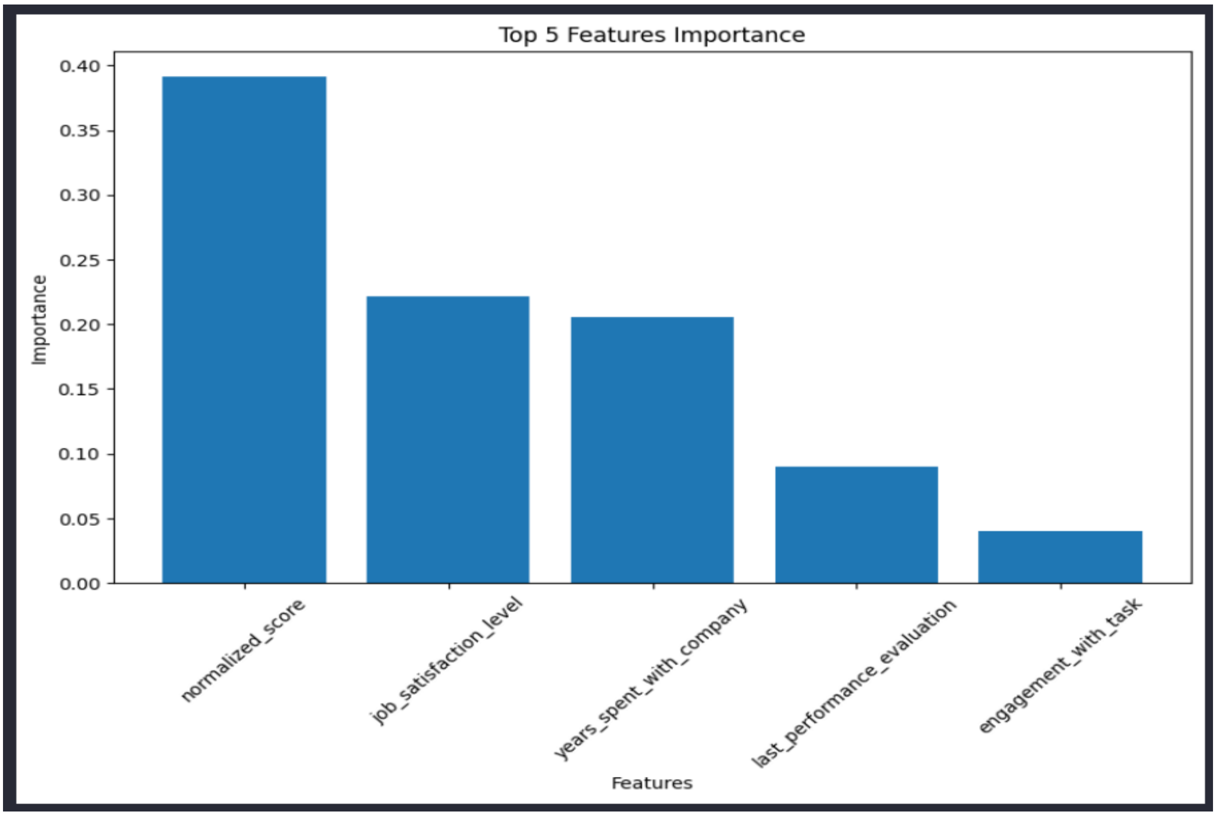
Model performance evaluation is conducted by generating the confusion matrix. From the figure, we can conclude that our model does a very good job at predicting whether an employee prefers to leave or stay with FAU Bank.



The overall accuracy of the model is 97% which is very good and other class level metrics are as follows

	precision	recall	f1-score	support
0	0.97	0.99	0.98	3428
1	0.97	0.90	0.94	1072
accuracy			0.97	4500
macro avg	0.97	0.95	0.96	4500
weighted avg	0.97	0.97	0.97	4500

By plotting the importance of each feature, we can find the most important features that contribute to employee turnover.



Here normalized_score is a new feature obtained from completed_projects and average_working_hours_monthly. So, from the feature importance graph, FAU Bank might want to focus on making sure that employees personal factors like – average working hours, job satisfaction level; years spent with company and last performance evaluations to increase employee retention.

References

1. Pfeffer, J. (2018). The overlooked essentials of employee well-being. McKinsey & Company.
2. Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta - Techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal, 29*(1), 6–42. https://doi.org/10.1111/isj.12169
3. Beaudry, A., & Pinsonneault, A. (2005). Understanding user responses to information technology: A coping model of user adaptation. *MIS Quarterly, 29*(3), 493-524. https://doi.org/10.2307/25148693
4. Bala, H., & Venkatesh, V. (2015). Adaptation to information technology: A holistic nomological network from implementation to job outcomes. *Management Science, 62*(1), 156-179. https://doi.org/10.1287/mnsc.2014.2111
5. Form, W. H. (1959). [Review of *Organizations.*, by J. G. March & H. A. Simon]. *Administrative Science Quarterly*, *4*(1), 129–131. https://doi.org/10.2307/2390654
6. Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. The Academy of Management Review, 19(1), 51-89. https://doi.org/10.2307/258835

Declaration of Academic Integrity at the Schöller Endowed Chair for Information Systems

I hereby certify that I have prepared the submitted work independently, and without the unauthorized assistance of third parties, as well as without the use of unauthorized aids. The work has not been submitted in the same or similar form to any other examination authority, nor has it been accepted by any other examination authority as part of an examination.

The passages in the work, which have been taken from other sources in terms of wording or meaning, are identified by indicating the origin. This also applies to drawings, sketches, picture representations and sources from the Internet.

I am aware that the use of artificial intelligence is permitted for work at the Schöller Endowed Chair of Information Systems, Digitalization in Business and Society (esp. to improve the text written by myself). However, the intellectual core of the respective work has been developed by me, and the scientific methods that are part of the work have been carried out by myself. Furthermore, I have transparently communicated the aids used in the work.

Violations of the above-mentioned rules are to be qualified as deception or attempted deception and lead to an assessment of the examination with "failed". Further sanctions are possible in the case of multiple or particularly drastic violations of the rules by the examination board.

G. Prabhu Nithin

Prabhu Nithin Gollapudi

Erlangen, 2024-07-15