“Should I Stay or Should I Go Now?” Applying Machine Learning to the FEVS to Predict Federal Employee Intent to Leave

Chuanfeng “Joy” Wu

Jennifer Ziegler

Georgetown University Data Science Certificate Cohort 27

September 24, 2022

Abstract

Using Machine Learning, Team Stay or Go analyzed five years of Federal Employee Viewpoint Survey data to identify the factors that predict an employee selecting “Leave” on the annual Intent to Leave question. After ingesting and wrangling survey data for 2016-2020, Team Stay or Go tested a 171-feature numeric dataframe using five machine learning model families. We used the model performances to reduce features to 17. Using Linear SVC, which had demonstrated the best combination of predictive ability and speed, we ran a sample of 1.8M rows and identified the top 13 features that predict an employee selecting something other than “Stay” in the FEVS survey. The project validated the use of the Global Satisfaction Index (GSI) and a portion of the Annual Employee Survey (AES), but not the Employee Engagement Index (EEI), to predict federal employee Intent to Leave.

Introduction and Inspiration

The idea to study the FEVS came from a classmate whose agency was interested in understanding more about why people leave the federal service, or conversely what makes people stay. We knew we wanted to work with a tabular data set, and we thought that the FEVS would give us a good opportunity to learn Python and other elements of machine learning while also answering a practical question. Early on, we riffed on the Clash Song "Should I Stay or Should I Go Now?" and this stuck as our group name, project theme, and ultimately the labels for our target variable in machine learning.

About the FEVS

Every year the Office of Personnel Management (OPM) invites all federal employees to express their sentiments about their current agency, role, work unit, supervisor, leaders, and other factors that affect their experience of work, using the FEVS. The survey is described as an "organizational climate survey" that "assesses how employees jointly experience the policies, practices, and procedures characteristic of their agency and its leadership" (OPM, 2020). For this project we studied the data from years 2016 to 2020. Table 1 summarizes agency participation, employee response rate, numbers of questions, and survey timelines for those five years. Management reports from the FEVS were typically published 4-5 months after each survey was administered. Notice that the 2020 survey was delayed from its typical spring-delivery-to-fall-reporting timeframe due to the Covid-19 pandemic.

Table 1: Summary of FEVS Timeline and Participation, 2016-2020; Source: OPM, 2016-2020

	2016	2017	2018	2019	2020
Agencies Participating	80	80	82	83	82
Employees Invited	889,590	1,068,151	1,473,870	1,443,152	1,410,610
Responses Received	407,789	486,105	598,003	615,395	624,800
Response Rate	45.8%	45.5%	40.6%	42.6%	44.3%
Workforce Questions	98	98	94	85	68
Demographic Questions	14	14	16	16	20
Total Questions	112	112	110	101	88
Dates Administered	April-May 2016	May 2017	April-May 2018	May 2019	Sept-Nov 2020
Date Published	Sept 20, 2016	Oct 12, 2017	Oct 15, 2018	Nov 7, 2019	April 26, 2021

Intent to Leave

To examine why employees might decide to leave or stay we focused on one question that is asked every year but that is not necessarily analyzed formally each year: "Are you considering leaving your organization within the next year, and if so, why?" These were asked in a slightly different ways in different years:

2016-2019: Are you considering leaving your organization within the next year, and if so, why?

A No

B Yes, to take another Federal job

C Yes, to take a job outside Federal Gov

D Other

2020: Are you considering leaving your organization within the next year, and if so, why? Today: (September-October 2020)

- A No
- B Yes, other
- C Yes, to take another job within the Federal Government
- D Yes, to take another job outside the Federal Government

Because the data are anonymized when reported it is not possible to track intention through to behavioral outcomes, nor how sentiments change from year to year. Our initial hypothesis was that by analyzing the FEVS data using Machine Learning, we could identify the factors that can predict someone answering "Yes, to take another job outside the Federal Government" to this question (i.e., "Leave").

FEVS "Indexes": The EEI, the GSI, and the NIQ

To our knowledge, only two analyses were completed on the Intent to Leave question during the period for this study. In 2016, a Special Report that accompanied the survey introduced the Employee Engagement Index (EEI), a model comprised of five "drivers" that were statistically identified from an analysis of 2013-2015 data: performance feedback, collaborative management, merit system principles, training and development, and work/life balance (OPM, 2016). Going forward, subsequent survey reports labeled the 15 questions in terms of their contributions to three "subfactors": Leaders Lead, Supervisor, and Intrinsic Work Experience (e.g., OPM, 2017). Reports in 2016 and 2018 linked the EEI with employee Intent to Leave: in 2016 EEI was reported to be 47% among those intending to leave vs. 72% among those intending to stay. A standalone infographic again called out the figures from the survey, with these admonitions:

> Engagement Matters! Engaged employees are: more innovative, more productive, more committed, more satisfied, and less likely to leave. Cost of turnover is high in terms of monetary and knowledge loss.
>
> EEI scores of employees who expressed intent to leave: 47%
> EEI score of employees who expressed intent to stay: 72%
> (source: OPM, 2016)

In 2018 the figures were reported to be 50% vs. 76% for leave vs. stay, respectively (OPM, 2018). Also in 2018, work-life balance was specifically linked to intention to leave. Management reports did not call out intention to leave again until 2021 (after the period for our analysis), at which time considering leaving was now tied to the Global Satisfaction Index (GSI). Specifically, the report highlights the managerial need to focus on GSI, claiming that "satisfied employees are more likely to stay in their jobs, reducing turnover" (OPM, 2021, Appendix A, p. 7).

Comparing Intent to Leave to the Quits Rate

Initially, we wondered to what extent turnover is a problem for the federal government. From the dataset we calculated that 3.6% of respondent responded “Leave” across the five years. To provide context for this number, we compared “Intent To Leave” to the Bureau of Labor Statistics “Quits” Rate published monthly at https://www.bls.gov/jlt/. The Quits Rate is the number of quits during the entire month as a percent of total employment (BLS, 2022). In December 2021, at the height of the Great Resignation, the BLS recorded the highest quits rate ever for all sectors at 2.9% (BLS, 2021). For that same month the federal employee Quits Rate was 0.7% (BLS, 2021). For the period we studied, federal employees quit their jobs at even lower rates from 2016 to 2020:

Table 2: Quits Rate for Time Period of the Study. Source: BLS, 2022

Month and Year	All Sectors	Federal Government Employees
May 2016	2.0%	0.4%
May 2017	2.2%	0.4%
May 2018	2.5%	0.5%
May 2019	2.3%	0.5%
Sept 2020	2.3%	0.6%

The data show that federal employees quit their jobs rates lower than people in other sectors. However, more employees seem to express Intent to Leave than actually quit. This may mean that the Intent to Leave question may be measuring something other than concrete plans to be in a different line of work 12 month hence. It may express a sentiment about the work experience at the moment of the survey, such as an indicator of workplace distress. Therefore, even though turnover might not necessarily be as large a problem for the federal government as for other industries, we still thought this a worthy question to study. However, given the half year lag to deploy, analyze, and report out the survey, and the aggregated nature of the results, we wondered if employee expression of Intent to Leave could represent an intervention opportunity for federal managers if they were to receive the data in a timelier way.

Project Goals and Approach

Our primary goal was to use machine learning to analyze the FEVS data to identify the factors that could predict an employee choosing “Leave” on the Intent to Leave question. A second project goal was to test the relationship between the Indexes and employee answers on the Intent To Leave question. Finally, if time allowed, we wanted to develop a web survey or app that could allow federal managers to ask a few key questions to predict future workforce availability or even have an opportunity to intervene depending on the results. There is some precedent for the latter from the private sector, such as one app that identified turnover shock and longevity as predictors of employee exit (Liu, 2019).

Data Ingestion and Wrangling

The OPM dramatically changed the FEVS survey from 2016 through 2020, which significantly complicated our work on this project. In 2017, the survey contained exactly the same text and number of questions as in 2016. But by 2016 the OPM had already started various processes to improve the FEVS including by developing various "Indexes" of questions that managers could use to measure specific constructs like Employee Engagement (the EEI), Global Satisfaction the (GSI), and feelings of inclusion (the New Inclusion Quotient, or NIQ). The OPM also had set a goal to reduce the congressionally mandated questions called the Annual Employee Survey (AES) from 40 to some smaller number (OPM, 2020).

By the time of its administration of the 2018 survey, the OPM had started to undertake these intended changes: text in some questions was changed to improve interpretation or understanding; and other questions were added, removed, or even combined. In 2018, 15 questions were different as compared to 2017. Then in 2019, 11 additional questions were changed or, in part to capture lingering effects of a partial government shutdown that had taken place earlier that year.

In 2020, as noted above, the Covid-19 pandemic delayed the administration of the survey. However, the pandemic also changed the nature of the questions asked. The "core" questions were pared back from 71 in 2019 to 38 in 2020, thus finalizing the streamlining of the AES to 16 questions; most of the Index questions were also maintained. This paring back helped to make room for Covid-19-specific questions relating to telework, safety, child / elder care, return to the workplace, etc. More detailed questions about motivations to retire or leave before and during the pandemic were also asked.

All told, even though about 100 questions are asked every year, when comparing the start of the data set to the end, only 41 of the questions that were asked in 2016 actually remained in 2020. (A note about demographics: The FEVS asked more nuanced questions about demographics as early as 2018, but likely due to low frequencies and inability to anonymize the data, these answers were aggregated to higher level categories in the publicly available data.)

Data Ingestion

We downloaded Public Release Data Files in CSV format for 2016-2020 from the OPM FEVS website at https://www.opm.gov/fevs/public-data-file/. The data files contained hundreds of thousands of instances (e.g., 615,000 in 2019) for 80+ columns (e.g., 85 in 2019). The actual FEVS questionnaire contained 90+ questions (e.g., 101 in 2019), which means that responses in the data files were "somewhat" already aggregated. This is evident by comparing Total Questions in Table 1 with Total variables in Table 3. That is, in some years, responses less than 10 were collapsed into other categories to de-identify the respondents.

Table 3: FEVS Data Set, 2016-2020

Public Data File	2016	2017	2018	2019	2020	Totals
File Size (MB)	75.5	83.9	111.8	128.8	177.4	577.4
Core Questions	71	71	71	71	38	
Demographics Questions	5	6	6	6	11	
Other Questions / Variables	4	3	3	8	87	
Total Variables (Columns)	80	80	80	85	136	188 unique

Total Respondents (Rows)	407,789	486,105	598,003	615,395	624,800	2,732,092
Expected Data Points (CxR)	32,623,120	38,888,400	47,840,240	52,308,575	84,972,800	256,633,135
Total Responses (non-Null Data Points)	30,282,319	36,293,100	44,890,253	51,345,364	76,952,601	239,763,637
Data Availability Rate	92.82%	93.33%	93.83%	98.16%	90.56%	93.43%
Cells with Null Data	2,340,801	2,595,300	2,949,987	963,211	8,020,199	16,869,498

Accompanying the datasets were Technical Reports that were helpful for understanding question changes and reporting changes from year to year. As one example, the phrase “considering leaving” was asked in the actual survey, whereas in the management reports the answers to this question were reported out as either “intent to leave” or “turnover,” depending on the year. Additionally, although the survey asked participants directly about their plans to retire, those answers were collapsed into “Other” when the data were reported out.

Kept All Questions. Early on, at the advice of our professors, we decided not to drop any data but rather to combine all questions from all years, even if this meant that we would be working with many null values. To keep track of our columns, we created a variable map in Excel where we noted which questions were asked each year, what sequence they appeared in for each year, and how they were numbered. We created a final table of all 188 questions, and labeled each question with a unique question ID. For example, the question, “I recommend my organization as a good place to work” was question #40 from 2016 to 2019, and question 17 in 2020. In our data set, this question is labeled “Q118” for all years. We also noted that it is one of the four questions included in the GSI. Because “Intent To Leave” was asked differently in different years we collected all the responses to this question into one column, and we labeled those responses “Stay,” “Retire,” “Transfer,” or “Leave.” This brought our column total to 189 going into early data analysis.

Figure 1: 189 Column Dataframe

```
In [6]:  FEVS5year.columns
Out[6]: Index(['response_id', 'year', 'agency_id', 'Q234', 'Q226', 'Q228', 'Q225',
               'Q236', 'Q237', 'Q102',
               ...
               'Q277', 'Q278', 'Q279', 'Q280', 'Q281', 'Q282', 'Q283', 'Q284', 'Q285',
               'StayorGo'],
              dtype='object', length=189)

In [7]:  FEVS5year.head()

Out[7]:
```

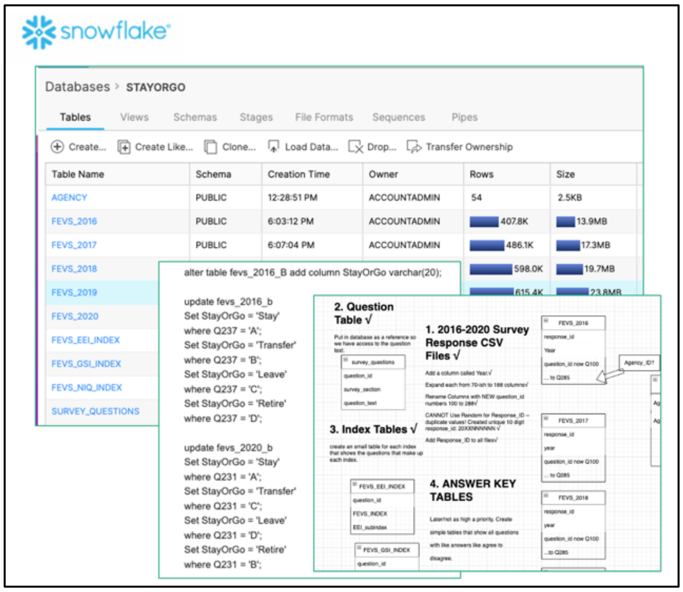
	response_id	year	agency_id	Q234	Q226	Q228	Q225	Q236	Q237	Q102	...	Q277	Q278	Q279	Q280	Q281	Q282	Q283	Q284	Q285	StayorGo
0	2016000002	2016	TR	TR93	A	B	B	A	A	5.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Stay
1	2016000003	2016	AF	AF1C	A	A	B	B	A	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Stay
2	2016000004	2016	TR	TRAD	A	A	B	B	A	5.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Stay
3	2016000005	2016	TR	TR93	A	A	B	B	D	3.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Retire
4	2016000006	2016	HE	HE09	B	B	B	B	A	5.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Stay

5 rows × 189 columns

Data Storage

Because we were working with a large dataset, we created an SQL database in Snowflake for WORM storage. Figure 2 shows our early table map, a sample script, and even a screenshot of the data after we loaded it into the system.

Figure 2: StayorGo SQL Database in Snowflake

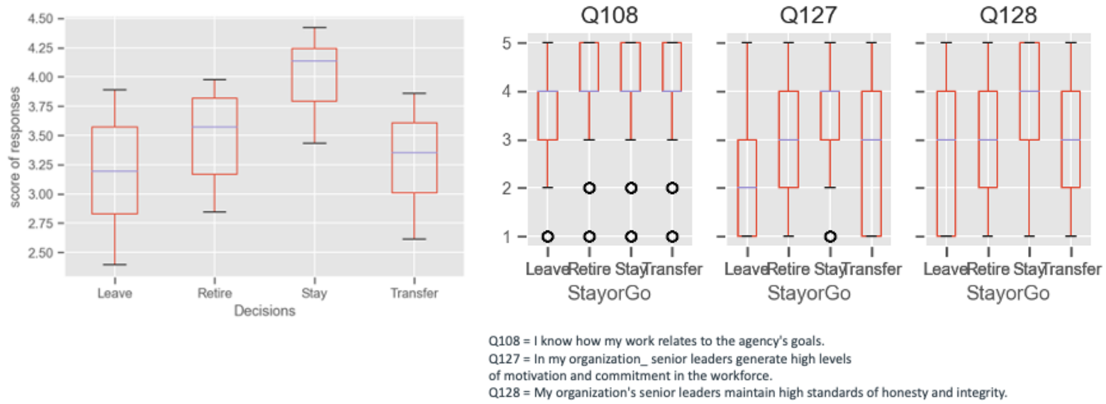


Snowflake is powerful because it combines storage, structure, and query capabilities. However, we experienced various administrative problems using the platform that were not easy to troubleshoot at first. In the interim, we used the Excel XLOOKUP function to rename the columns for each year in the actual CSV files. After we standardized the columns in this way, we found that it was easy enough to work with the CSV files in our Jupyter Notebooks. So, while setting up the Snowflake database was a great experience to learn about SQL and the Snowflake platform, we found that we didn't really use the database after we set it up.

Exploratory Data Analysis

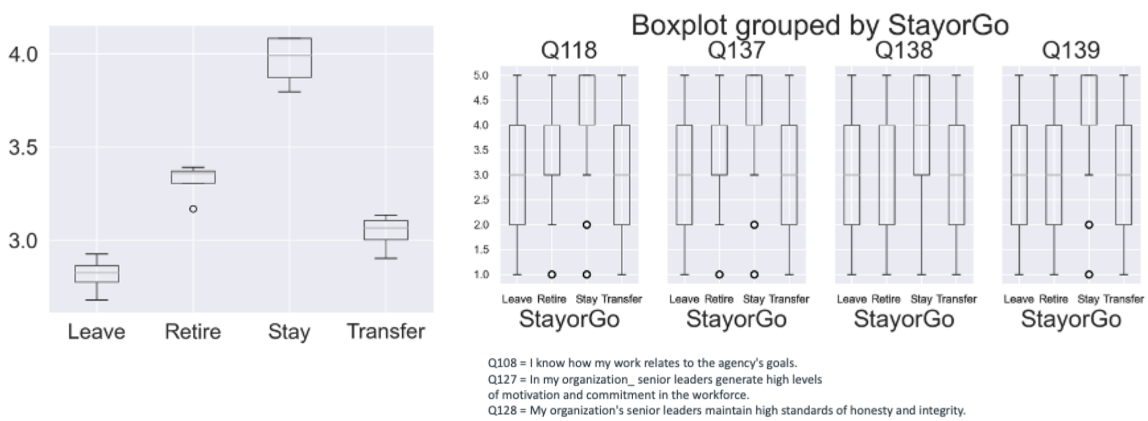
In our early exploratory data analysis, we examined the relationship between the EEI and the Leave, and the GSI and Leave. On average, employees who indicated “Leave” scored lower on EEI questions than those indicating “Stay.” We took a closer look at the three questions that appear in Figure 3.

Figure 3: Correlation between Leave and EEI Score and Selected Questions in early Exploratory Data Analysis.



Next, we analyzed the four questions related to the Global Satisfaction Index (GSI) for all years of the survey, looking for their correlation with Intent to Leave or Stay. In the boxplots in Figure 4, it is evident that those answering "Leave" rated answers to the GSI questions lowest of all. Additionally, Leave and Transfer responded similarly on GSI questions, but Stay responses looked dramatically different.

Figure 4: Correlation between Leave and GSI Score and Selected Questions in early EDA



Data Wrangling

The original dataset contained some numeric data from 5-item Likert Scale questions. However, most columns consisted of other data types that required transformation.

"Almost Numeric" Data. Most of the Likert scale questions contained the option "Do Not Know" which introduced the letter "X" into our data. This automatically caused Pandas to read these fields as strings. To address this, we coerced these fields into numeric form using pandas. After that, Xs became NaNs (Not a Number), and these columns were recognized as numeric. See Figure 5 as one example.

Figure 5: Using Pandas to Coerce Strings to Numbers

Q119=

My talents are used well in the workplace.		
	5	Strongly Agree
	4	Agree
	3	Neither Agree nor Disagree
	2	Disagree
	1	Strongly Disagree
	X	Do Not Know

```
In [22]: print(FEVS5year['Q119'].unique())

['5' '2' '4' '3' 'X' nan '1']
```

```
In [35]: FEVS5year3[['Q103','Q104','Q106','Q107','Q113','Q118',
                 'Q119','Q126','Q133','Q136','Q137','Q138','Q139',]]= FEVS5year3[
                ['Q103','Q104','Q106','Q107','Q113','Q118','Q119',
                  'Q126','Q133','Q136','Q137','Q138','Q139',]].apply(pd.to_numeric,errors='coerce')
         FEVS5year3.Q119.unique()

Out[35]: array([ 5.,  2.,  4.,  3., nan,  1.])
```

Encoding Demographic Data. Answers to demographic questions were captured as categorical data with labels like A or B. We used special encoding to convert these categorical data to binary. For example, Figure 6 depicts how Question 226 labeled answers about supervisory status as A or B. In Pandas, after calling "get_dummies" on our dataframe, question 226 became columns 226_A and 226_B with zeros and ones to indicate present or not present.

Figure 6: Using Pandas to Convert Categorical Data to Binary

Q226=

What is your supervisory status?		
	A	Non-Supervisor/Team Leader
	B	Supervisor/Manager/Executive

```
In [26]: FEVS5year3 = pd.get_dummies(FEVS5year2, columns = ['Q226','Q227','Q228','
         print(FEVS5year3)

        agency_id  Q103  Q104 Q106 Q107 Q113  Q118 Q119  Q126 Q133  ...  \
0              TR   2.0   4.0    4    4    2   5.0    5   4.0    3  ...
1              AF   3.0   3.0    4    2    2   2.0    2   3.0    X  ...
2              TR   5.0   5.0    5  NaN    4   5.0    4   4.0  NaN  ...
3              TR   4.0   4.0    4    4    3   4.0    3   5.0    3  ...
4              HE   5.0   5.0    4    5    4   5.0    4   5.0    5  ...
...           ...   ...   ...  ...  ...  ...   ...  ...   ...  ...  ...
624795         HS   5.0   4.0    2    4    2   3.0    2   4.0    2  ...
624796         HS   5.0   4.0    5    5    4   5.0    5   5.0    4  ...
624797         HS   2.0   3.0    3    4    4   1.0    1   5.0    3  ...
624798         HS   3.0   4.0    4    3    3   4.0    3   5.0    4  ...
624799         HS   5.0   5.0    5    5    2   5.0    4   4.0    5  ...

        Q226_A  Q226_B  Q227_A  Q227_B Q227_C  Q228_A  Q228_B  Q235_A  Q235_B
0            1       0       0       0      0       0       1       0       0
1            1       0       0       0      0       1       0       0       0
2            1       0       0       0      0       1       0       0       0
3            1       0       0       0      0       1       0       0       0
4            0       1       0       0      0       0       1       0       0
```

Many null values existed in our data set due to keeping all questions from all years and due to coercing strings to numbers. However, before going into machine learning we needed to transform these NaNs (Not a Number) to numbers in some way. For Likert scale questions that contained "Do Not Know" we decided to impute NaN using the column mean for each question. Figure 7 depicts an example for Question 107 which was "my talents are used well in the workplace." After we coerced Question 107's Xs to NaN's, we then imputed the column mean (3.5) such that all NaNs in that column were replaced with that mean value.

Figure 7: Using a For-Loop to Impute NaN in Selected Columns

Q107=

My talents are used well in the workplace.		
	5	Strongly Agree
	4	Agree
	3	Neither Agree nor Disagree
	2	Disagree
	1	Strongly Disagree
	X	Do Not Know

```
In [49]:  FEVS5yearT2.Q107.unique()
Out[49]: array([ 4.,  2., nan,  5.,  3.,  1.])

In [54]:  impute = {}

          for col in cols:
              impute[col] = FEVS5yearT2[col].mean()

In [55]:  FEVS5yearT2.fillna(value=impute, inplace=True)

In [56]:  FEVS5yearT2.Q107.unique()

Out[56]: array([4.         , 2.         , 3.52379154, 5.         , 3.         ,
                1.         ])
```

For our target variable, we noticed a class imbalance for “Leave” where 3.6% was lowest frequency of all the answer options (even less frequent than missing values). As shown in Figure 8, we transformed Stay or Go to a binary with two options: Any option other than Stay was put into the “Go” category, including “Retire,” “Transfer,” and “Leave.” We filled the NaN values proportionally, trying to maintain the 64% proportion for Stay in the final frequencies for Stay and Go.

Figure 8: Transforming Stay or Go to Binary and Filling NaN Proportionally.

```
In [8]:  # determine the number of NaN cases in Stay or Go
         print(FEVS5year['StayorGo'].unique())
         FEVS5year['StayorGo'].value_counts(dropna=False)

         ['Stay' 'Retire' 'Transfer' nan 'Leave']

Out[8]:  Stay        1736122
         Transfer     442466
         Retire       278698
         NaN          176148
         Leave         98658
         Name: StayorGo, dtype: int64

         Stay = 0.6354551750087479
         Transfer = 0.1619513544931869
         Retire = 0.10200900994549232
         Leave = 0.036110789826989724
         NaN = 0.06447367072558317
         --
         Stay = 0.6354551750087479
         any_not_stay = 0.30007115426566894
         NaN = 0.06447367072558317
```

```
In [10]:  df.StayorGo.unique()
Out[10]: array(['Stay', 'Retire', 'Transfer', 'Leave', 'Go'], dtype=object)

In [11]:  df['StayorGo'].value_counts(dropna=False)
Out[11]: Stay        1855770
         Transfer     442466
         Retire       278698
         Leave         98658
         Go            56500
         Name: StayorGo, dtype: int64

In [12]:  df=df.replace(to_replace='Retire', value='Go')
          df=df.replace(to_replace='Transfer', value='Go')
          df=df.replace(to_replace='Leave', value='Go')

In [13]:  df['StayorGo'].value_counts(dropna=False)
Out[13]: Stay    1855770
         Go       876322
         Name: StayorGo, dtype: int64
```

<u>Additional Dropped Columns</u>. After label encoding our demographic variables, our dataframe had grown to over 195 columns. We looked very closely at the questions and decided to discard an additional 29 questions that did not seem generalizable to the entire data set. These were questions that were very idiosyncratic for a specific year such as employee experience with a certain HR policy during the shutdown, for example. After we dropped those 29 columns, we entered machine learning with a dataframe that had 171 features, plus our target column of Stay or Go.

Modeling and Visualization

Classification model (i.e., supervised learning) families were most appropriate for our data because we were interested in developing the simplest, most explanatory model that will predict an employee's expressed intention to stay with or leave their federal agency within the year. For modeling, our instances were 2.7M employee responses, with 171 variables used to predict an employee's expressed intent to Stay or Go.

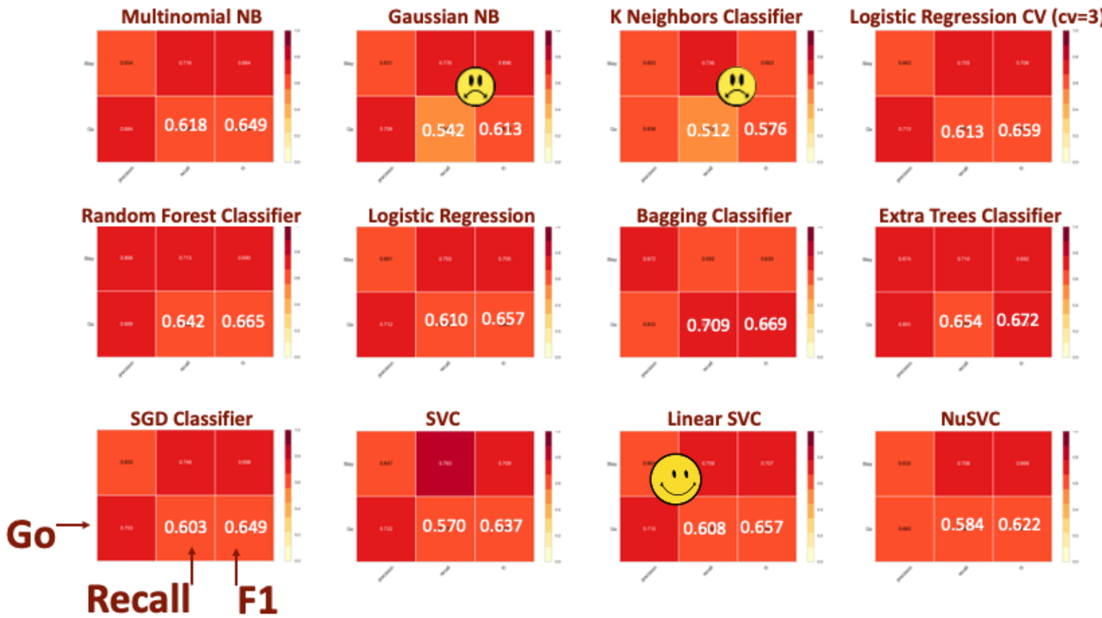
We evaluated five model families with Binary Classification Algorithms:

- Naive Bayes
  - Multinomial NB ()
  - Gaussian NB ()
- Support Vector Classifier
  - SVC (gamma='auto')
  - Nu SVC (gamma='auto')
  - Linear SVC ()
- Logistic Regression
  - Logistic Regression (solver='lbfgs')
  - Logistic Regression CV (cv=3)
- K-Nearest Neighbors
  - K Neighbors Classifier ()
- Ensemble Classifiers
  - SGD Classifier (max_iter = 100, tol=1e-3)
  - Bagging Classifier ()
  - Extra Trees Classifier (n_estimators = 100)
  - Random Forest Classifier (n_estimators = 100)

With 171 features and 2.7M rows, we ran into challenges with computing power. Thus, for initial model exploration, we sampled 20,000 cases from the data (10,000 rows each for Stay and for Go). We set the random seed to 100 so that we could maintain consistent results.

We examined Classification Reports to compare Recall Scores for "Go" as well as overall F1 Scores. Many of the model results were similar. As noted in Figure 9, some of the Ensemble models plus Linear SVC scored the best, while Gaussian Naïve Bayes and KNeighbors Classifier performed the worst.

Figure 9: Classification Reports for Various ML Models; n=20,000



Initially, we experimented with both the larger (171 features) and a smaller data set (i.e., just the questions in the indexes that were asked all five years), but we found evidence to engage in feature analysis with the larger data set, so we discarded the smaller set. We then compared the model performance to one another using classification reports, feature importances (where available), the class prediction error reports, cross-validation scores, and confusion matrices. (See Figures 10 and 11.)

Figure 10: ClassPredictionError () and StratifiedKFold(K=12) scoring='f1_weighted' Reports for 20,000 Cases

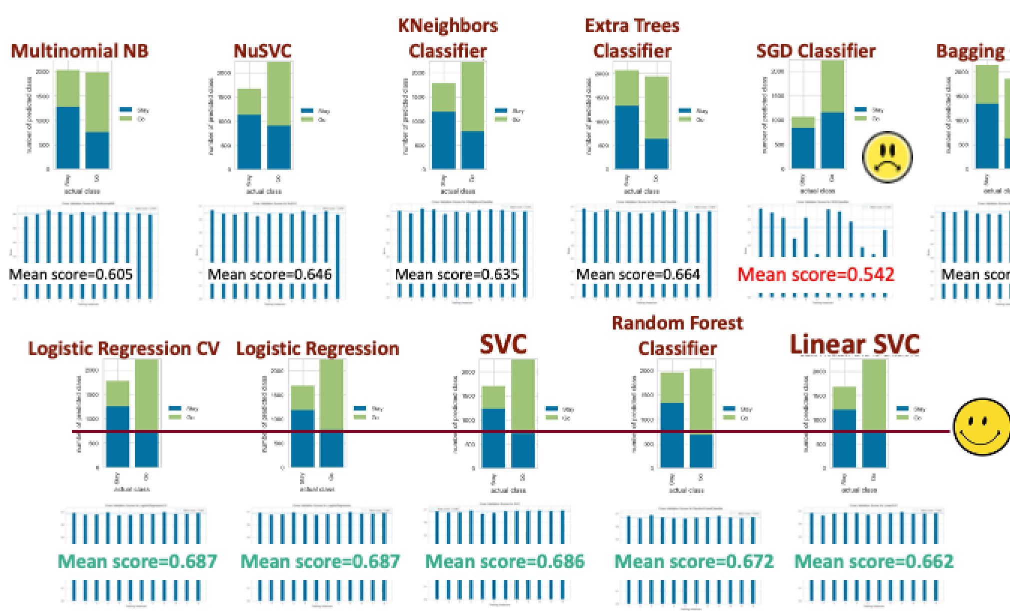
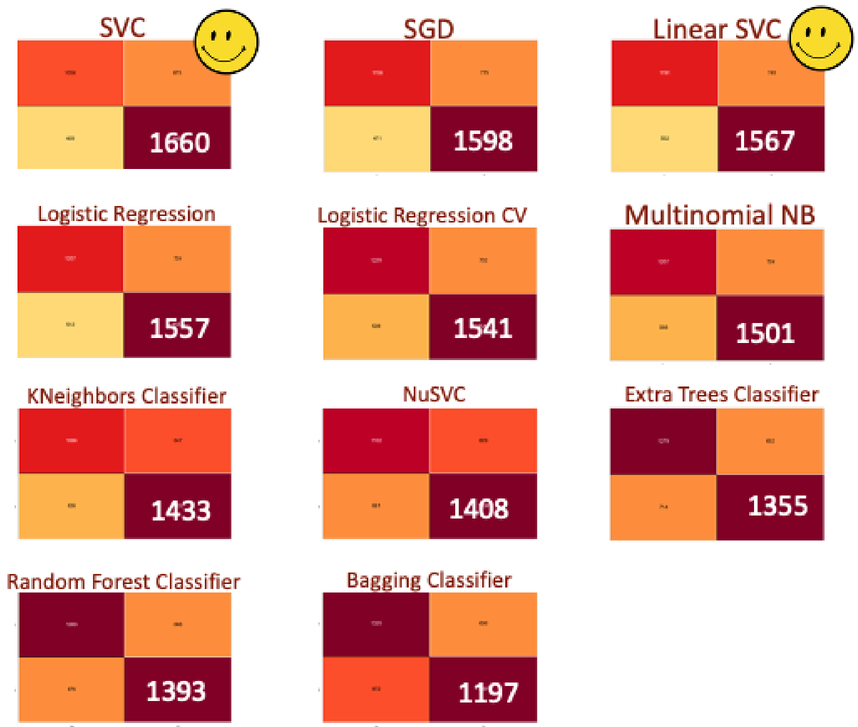


Figure 11: Confusion Matrices for ML Models (n=20,000)



We used these performances to engage in recursive feature elimination. That is, we kept features that stronger models deemed more important, and we iteratively removed features that were not deemed as important by the strongest models. We iterated from Large to Medium to Tiny to Tiny_2 data sets. Figure 12 shows the final 17 features and their top 23 importances from various models.

Figure 12: Final 17 Features in Tiny_2 Data Set (23 Columns Due to Label Encoding)

Label	question_id	Linear SVC	SGD Classifier	Logistic Regression	Logistic RegressionCV	Extra Trees	Random Forest
SatisfJob	Q137	1	1	1	1	1	4
TalentsUsed	Q107	4	6	4	4	4	5
RecommendOrg	Q118	2	4	2	2	7	8
SatisfOrg	Q139	6	5	7	7	12	3
SatisfPay	Q138	7	7	6	6	10	10
SurveyBetter	Q119	11	9	16	16	2	1
SrLSupportWL	Q133	16	21	13	13	6	6
PerfDiffRecog	Q113	18	13	19	19	3	2
GoodJobBySup	Q126	10	20	10	10	11	12
EducBach	Q235_B	8	12	8	8	15	15
SatisfRecog	Q136	14	15	14	14	8	11
WorkloadReas	Q106	17	11	18	18	5	7
PersAccomp	Q104	15	16	11	11	13	13
NonSupervisor	Q226_A	3	2	3	3	22	22
Female	Q228_B	9	14	9	9	18	17
Supervisor	Q226_B	5	3	5	5	23	23
EducMoreBach	Q235_C	12	8	15	15	14	14
EducLessBach	Q235_A	13	10	12	12	16	19
EncourBetter	Q103	21	18	21	21	9	9
Male	Q228_A	20	19	20	20	17	16
11to20Yr	Q227_B	19	17	17	17	19	20
10orLessYr	Q227_A	22	22	22	22	20	18
21orMoreYr	Q227_C	23	23	23	23	21	21
	StayOrGo						
	agency_id						

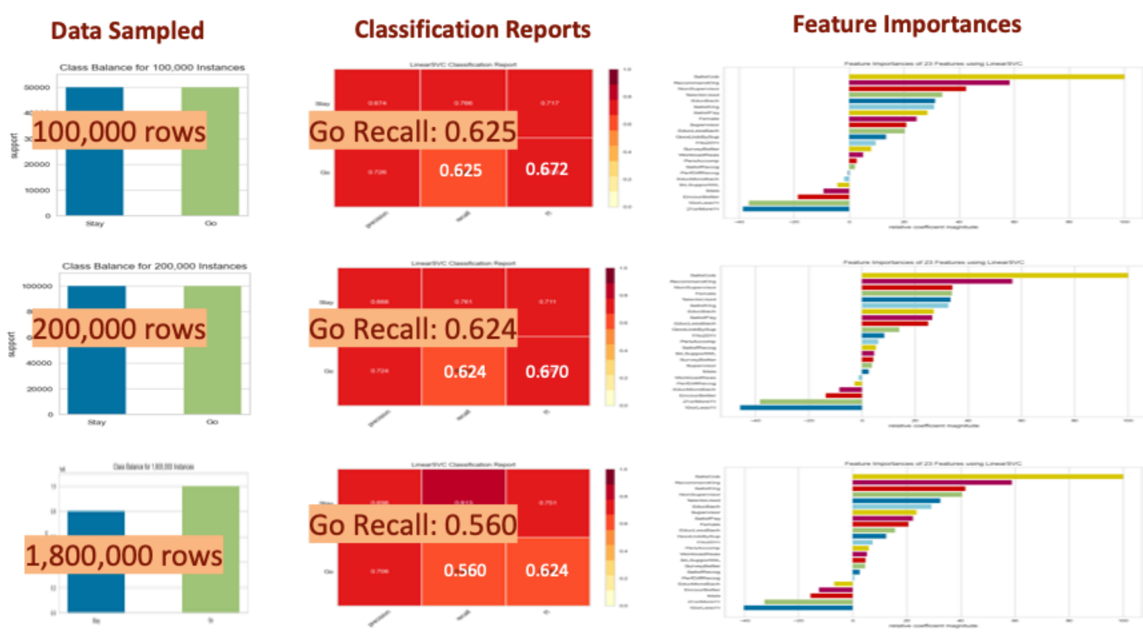
Model Testing

Two models had performed the strongest: Linear SVC and SVC. As we learned in class, Support Vector Classifier works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes (Bengfort, 2022). Linear SVC allowed for introduction of the most additional rows without a significant increase in computation time, so we chose to move forward with Linear SVC.

Linear SVC Performance

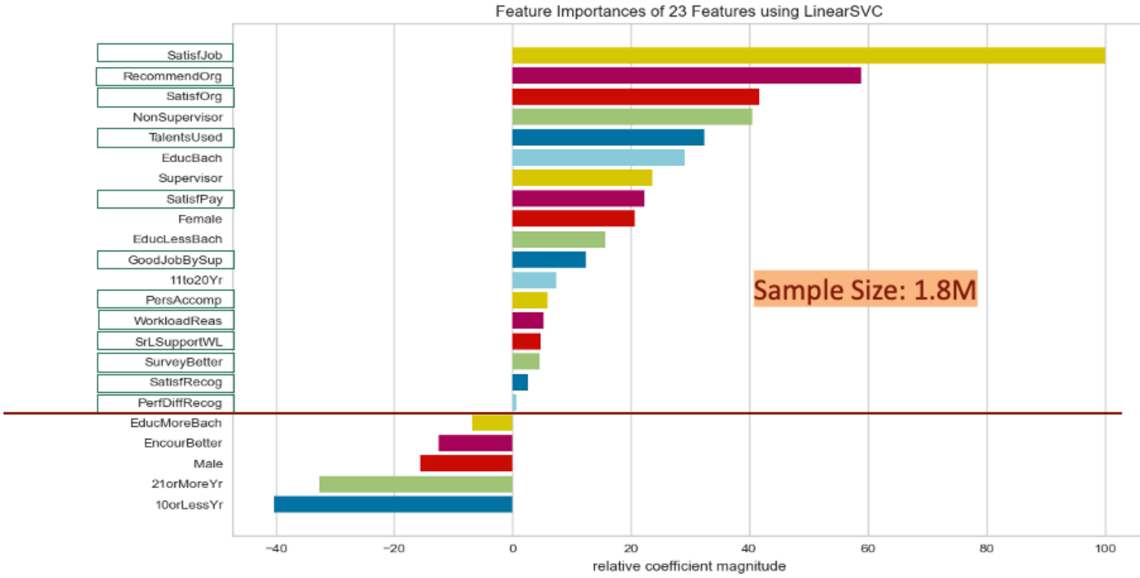
As noted in Figure 13, we were able to sample increasing numbers of rows (100K, 200K, and 1.8M). Even with the larger sample sizes we did not see a dramatic improvement in “Go” Recall Scores or F1 Scores. Indeed, when we replicated the class imbalance for 1.8M rows, both scores went down slightly to 0.560 for Go Recall and to 0.624 for F1 score. Additionally, increasing the sample size did not dramatically change the Feature Importances reported for Linear SVC.

Figure 13: Class Balance, Classification Reports, and Feature Importances for Three Sample Sizes: n=100K, n=200K, n=1.8M



A closer examination of Feature Importances (Figure 14) shows that Linear SVC points to 12 substantive features and six demographic options that are predictive of Stay or Go. We chose to focus on the substantive features in our additional data analysis and discussion.

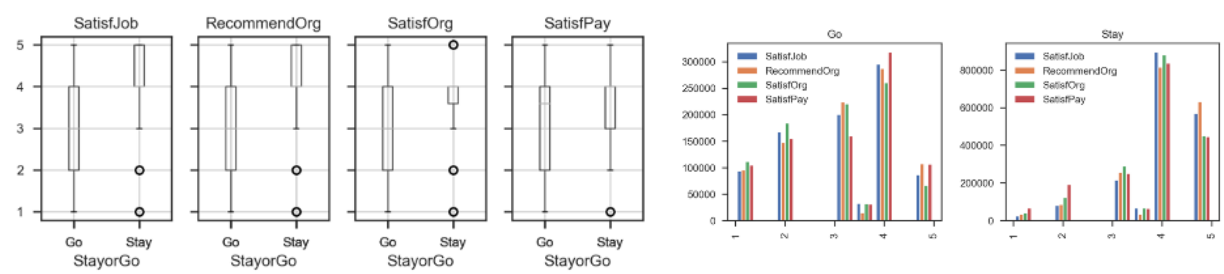
Figure 14: Feature Importances from Linear SVC (n=1.8M rows)



Additional EDA

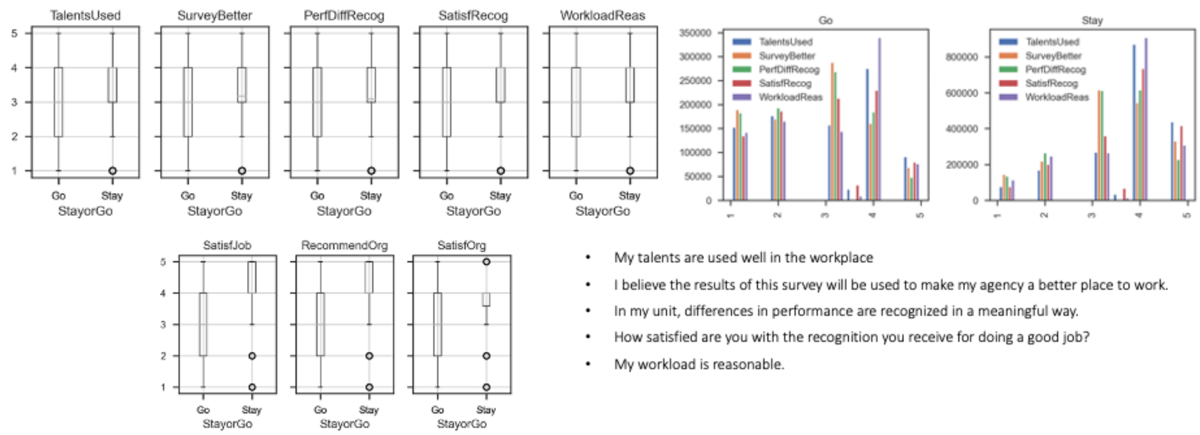
At the top of our list of Feature Importances were the four questions that make up the global satisfaction index: job satisfaction, satisfaction with one's pay, satisfaction with one's organization, and whether an employee recommends their agency as a good place to work. Returning to the entire data set (n=2.7M), and as illustrated in Figure we noted that those who answered more 1s, 2s, and 3s, proportionally, on these questions are those who are more likely to express intent to Go.

Figure 15: A Comparison of GSI Questions and Stay or Go (n=2.7M)



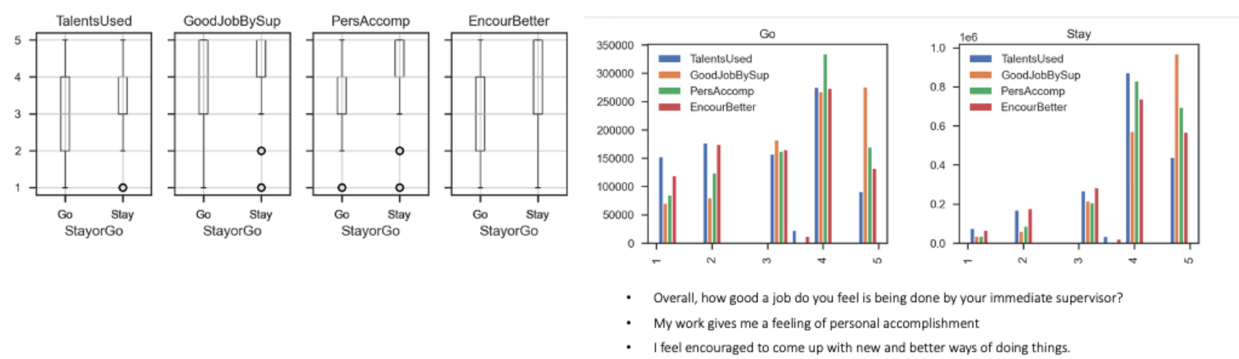
Eight out of the 16 Annual Employee Survey (AES) questions mandated by Congress contributed to our ML model in predicting an employee's intent to Go (See Figure 16). Three of these overlap with the GSI, but additional unique ones include questions about talents, recognition, workload, and even whether this survey will have an impact.

Figure 16: A Comparison of Eight of the AES Questions with Stay or Go (n=2.7M)



Finally, individuals who rated certain other factors low are also likely to indicate Go. As noted in Figure 17, three additional questions appeared in our top features, having to do with one's rating of their supervisor, feeling accomplished at work, and being encouraged to innovate. Again, employees who selected Stay and those who indicated Go may have answered all five options, but those who expressed intent to Go selected 1s, 2s, and 3s more often than those who indicated Stay.

Figure 17: Other Important Features Compared to Stay or Go (n=2.7M)



Despite their prominence in the final Feature Analysis list, supervisor status, gender, and education did not seem to contribute to whether an employee will select Stay or Go.

Conclusions

We succeeded with our primary project goal: By analyzing five years' of FEVS data using machine learning, we were able to identify the most important factors that can predict an employee's Intent to Leave. The top 13 substantive factors are listed in Figure 18, with a key to the Index(es) in which each question appears.

Figure 18: Top Features that Predict Employee Intention to Stay or Go

AES	EEI	GSI	Rank	Question Text
√		√	1	Considering everything, how satisfied are you with your job? 5 = Very Satisfied ... 1 = Very Dissatisfied
√		√	2	I recommend my organization as a good place to work. 5 = Strongly Agree ... 1 = Strongly Disagree
√		√	3	Considering everything, how satisfied are you with your organization? 5 = Very Satisfied ... 1 = Very Dissatisfied
√	√		5	My talents are used well in the workplace. 5 = Strongly Agree ... 1 = Strongly Disagree; X = Do Not Know
		√	8	Considering everything, how satisfied are you with your pay? 5 = Very Satisfied ... 1 = Very Dissatisfied
	√		11	Overall, how good a job do you feel is being done by your immediate supervisor? 5 = Very Good ... 1 = Very Poor
	√		13	My work gives me a feeling of personal accomplishment. 5 = Strongly Agree ... 1 = Strongly Disagree
√			14	My workload is reasonable. 5 = Strongly Agree ... 1 = Strongly Disagree; X = Do Not Know
			15	Senior leaders demonstrate support for Work/Life programs. 5 = Strongly Agree ... 1 = Strongly Disagree; X = Do Not Know
√			16	I believe the results of this survey will be used to make my agency a better place to work. 5 = Strongly Agree ... 1 = Strongly Disagree; X = Do Not Know
√			17	How satisfied are you with the recognition you receive for doing a good job? 5 = Very Satisfied ... 1 = Very Dissatisfied
√			18	In my work unit, differences in performance are recognized in a meaningful way. 5 = Strongly Agree ... 1 = Strongly Disagree; X = Do Not Know
	√		20	I feel encouraged to come up with new and better ways of doing things. 5 = Strongly Agree ... 1 = Strongly Disagree
8	4	4		
50%	27%	100%		Percent of index captured in top 13 substantive features

For our secondary goal, our results validated the Global Satisfaction Index as a good predictor of Intent to Leave, as 100% of those questions showed up in the top five feature importances. Additionally, half of the questions mandated to be included in the Annual Employee Survey also appear on the list as good predictors of Intent to Leave. On the other hand, we found no support for the Employee Engagement Index's ability to predict Intent to Leave. Only four of the fifteen questions in that index appeared in our top 13. As for developing a web survey or an app, while we ran out of time before being able to entertain the possibility, we still think it would be a good idea to try.

Practical Application

Based on these results, we recommend that when trying to predict employee Intent to Leave, OPM should focus on the Global Satisfaction Index (GSI) and the answers to all four of those questions, as well as answers to eight out of the 16 Annual Employee Survey questions as noted in Figure 18. We would also caution against using the Employee Engagement Index (EEI) as a tool for predicting Intent To Leave, as only four of the EEI questions made it into our top 13. As noted earlier, the Employee Engagement Index is comprised of five questions each related to Leaders, Supervisors, and Intrinsic Work Environment (IWE) factors. Three of the five IWE, only one of the Supervisor questions, and none of the Leader questions appeared in our final Feature Importances. This suggests that mainly Individual Workplace factors, and only one's impression of one's immediate Supervisor, may have an impact on an employee expressing Intent to Leave. Employee Engagement might still be good to measure for other reasons (e.g., checking employee perceptions of leader accountability at the agency level), but not necessarily for predicting Intent to Leave.

Lessons Learned and Future Work

Limitations

As noted above, due to challenges with processing such a large data set, we modeled only sample data. However, we also noted that scores did not necessarily improve as we increased the sample size. Second, by collapsing "Retire" and "Transfer" together with "Leave" in "Go," we may have diluted our initial focus on employees intending to exit the system to work outside the federal government. We believe studying the latter is still a worthy goal, but it would be important to address the class imbalance for machine learning to provide accurate results. Also, by combining all “Not Stay” into one category, the BLS "Quits" rate might no longer be the comparable metric. It may be more accurate to compare “Go” to the BLS measure of “Total Separations,” for example. Finally, because we did not use the weighting variables that were included in the codebooks that adjusted for agency size and differences in response rates, our conclusions are only generalizable to the federal government as a whole and not necessarily to individual agencies.

Future Opportunities

Anyone continuing with this data set in the future could incorporate the weighting variable to break out results by agency. One could also take a closer look at specific questions about leaving that were asked with respect to the Shutdown in 2019 and the Pandemic in 2020.

Lessons Learned

As a team we learned a few lessons during the execution of this project. First, there were more demographic questions asked across the years, and more answer options given, but these were aggregated in the data set to protect privacy. In the future we might try harder in the future to obtain the raw data set as it would have enriched our analysis to work with more nuanced demographic data. Of course, we would take care to protect privacy in any reporting. Second, while we appreciated learning about SQL (and Snowflake in particular), we look forward to working on future projects that require more extensive querying and analysis using the database.

Regarding lessons about data science, we learned that data wrangling takes longer than we expected, data balance impacts Machine Learning model performance, feature selection and analysis are iterative with Machine Learning, and multiple factors need to be considered (e.g., F1 scores, cross validation scores, confusion matrices, etc.) when evaluating a Machine Learning model. Fortunately, we also learned that data science takes place in the context of community: we learned to Google our questions and ask for help of instructors and colleagues. We sincerely appreciate the many people who assisted us with this project.

References

Bengfort, B. (2022). Machine Learning with Python: Session 5. Presented in XBUS-505 Data Analysis II: Machine Learning.

Liu, J. (2019, Sep 10). This algorithm can predict when workers are about to quit—here's how. CNBC. https://www.cnbc.com/2019/09/10/this-algorithm-can-predict-when-workers-are-about-to-quitheres-how.html

U.S. Bureau of Labor Statistics (2021, Feb 1). Job openings and labor turnover – December 2021. https://www.bls.gov/news.release/archives/jolts_02012022.pdf.

U.S. Bureau of Labor Statistics (2022). Job openings and labor turnover archived news releases. https://www.bls.gov/bls/news-release/jolts.htm.

U.S. Office of Personnel Management. (2016). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.

U.S. Office of Personnel Management. (2017). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.

U.S. Office of Personnel Management. (2018). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.

U.S. Office of Personnel Management. (2019). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.

U.S. Office of Personnel Management. (2020). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.

U.S. Office of Personnel Management. (2021). Federal employee viewpoint survey: Governmentwide management report. https://www.opm.gov/fevs/reports/governmentwide-reports.