Empower U.S. military members, veterans, and military spouses to succeed in the civilian workforce: Using an Analytical approach

Teradata Data Challenge

Team

Anjali Khushalani
MS Data Science & Business Analytics

Dikshya Mohanty
MS Data Science & Business Analytics

Ramkumar Nottath
MS Data Science & Business Analytics

Tarun Sharma
MS Data Science & Business Analytics

Varun Varia
MS Computer Science

January 2019

National Unemployment Rate
4%

Non-Veteran Unemployment Rate
4.3%

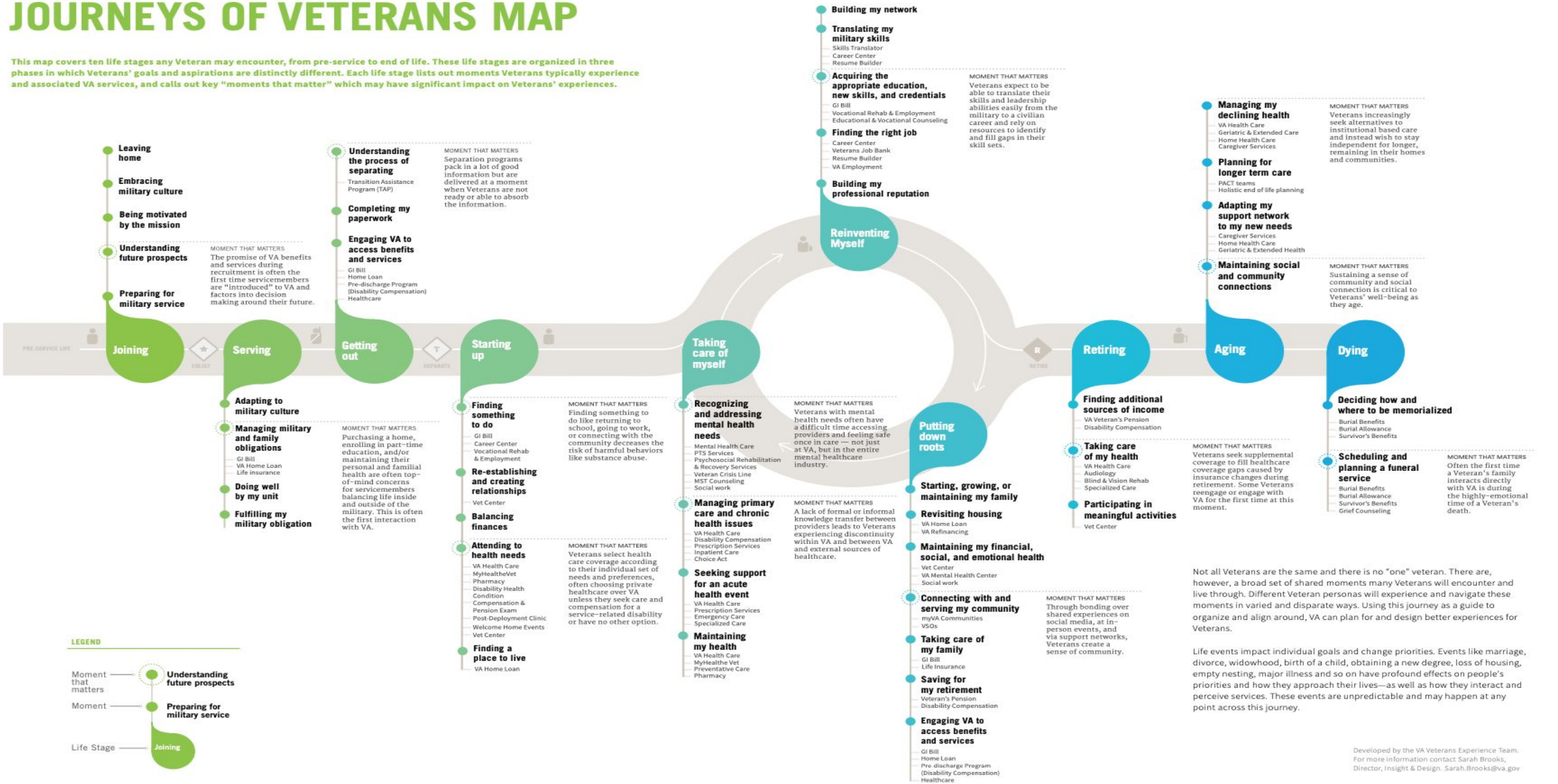
Veteran Unemployment Rate
3.7%

Women Veteran Unemployment Rate
2.7%

https://www.dol.gov/vets/newsletter/

JOURNEYS OF VETERANS MAP

This map covers ten life stages any Veteran may encounter, from pre-service to end of life. These life stages are organized in three phases in which Veterans' goals and aspirations are distinctly different. Each life stage lists out moments Veterans typically experience and associated VA services, and calls out key "moments that matter" which may have significant impact on Veterans' experiences.



Not all Veterans are the same and there is no "one" veteran. There are, however, a broad set of shared moments many Veterans will encounter and live through. Different Veteran personas will experience and navigate these moments in varied and disparate ways. Using this journey as a guide to organize and align around, VA can plan for and design better experiences for Veterans.

Life events impact individual goals and change priorities. Events like marriage, divorce, widowhood, birth of a child, obtaining a new degree, loss of housing, empty nesting, major illness and so on have profound effects on people's priorities and how they approach their lives—as well as how they interact and perceive services. These events are unpredictable and may happen at any point across this journey.

Developed by the VA Veterans Experience Team.
For more information contact Sarah Brooks,
Director, Insight & Design. Sarah.Brooks@va.gov

SERVING IN AND SEPARATING FROM THE MILITARY | **LIVING CIVILIAN LIFE** | **RETIRING AND AGING**

Motivation

- **United States is home to millions of Veteran military families**
- **Veterans offer a unique set of skills, experiences and leadership abilities**
  - **Translating these skills to the civilian workforce and marketing themselves to employers plays a major role**
- **HHUSA is an organization which is helping them to achieve stability in civilian life**
- ***U.S. Department of Labor estimates that the military trains people in skills applicable in at least 962 civilian occupations****

http://www.ncsl.org/Portals/1/Documents/Labor/Licensing/vetsFamilies_v03_web.pdf

How to get a veteran hired in the shortest span?

Out of the several challenging business questions that HHUSA provided, our focus was on the primary function of HHUSA which is nothing but getting a veteran hired

- **Is there any relationship between the amount of time spent working with individual clients (time to complete an assessment, time to complete resume, # of logged activities, etc.) and how quickly are they employed?**
- **Is there a relationship between a client's demographic profile and getting hired?**

What are we aiming to achieve of this reserach?

- **To be able to provide targeted training to the unemployed veterans**
- **Improved hiring profiles can be established to target veteran with specific configurations of life, military, and work experience**
- **Tapping the pockets of filled with military talent and unique skills**
  - **This solution solves the moral imperative and becomes a competitive business advantage**

Dataset

Description
• **Data is provided by Hire Heroes USA** • **Non Profit Organization** • **Empowers U.S. military members, veterans, spouses** • **Goal is to support military personnel to succeed in civilian workforce**

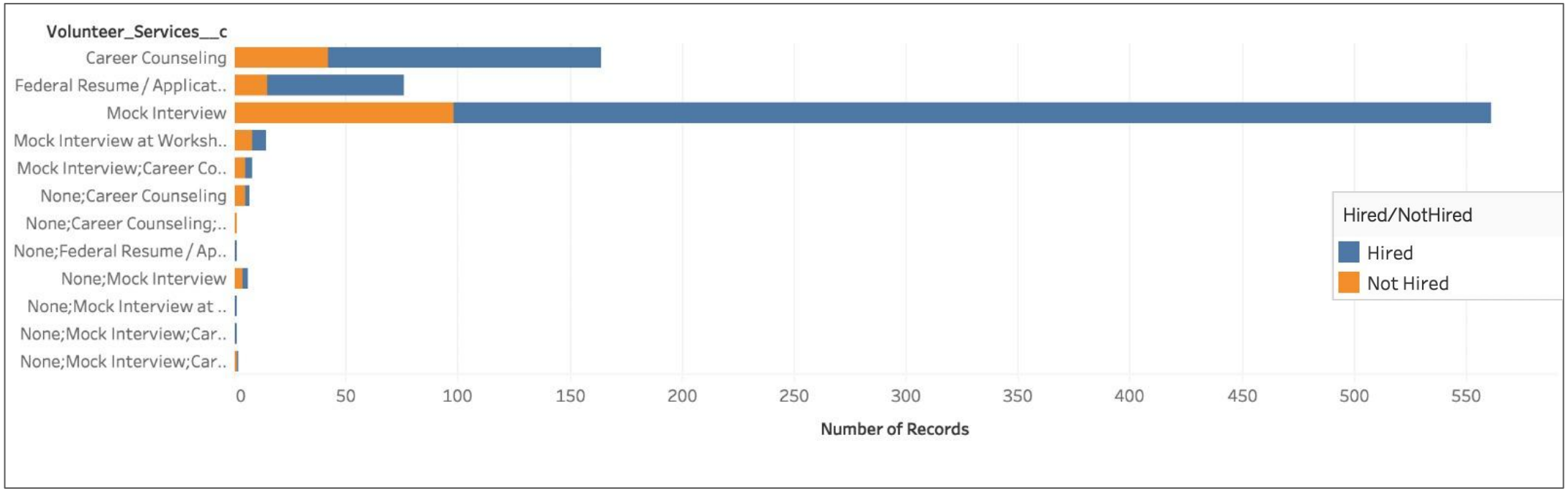
Datasets

• **Salesforce CRM structure**
• **Data provided includes 13 spreadsheets**
• **To answer our questions we focused on:**
  ○ **Contacts**
  ○ **Activities**
  ○ **Hired Information**

Summary

Variable	Number of Missing Values	% of Missing Values
Gender	1046	7.60
Last Rank	12685	92.26
Race	10750	78.19
Additional Service Needs	13606	98.96
Status	1075	7.81

How Volunteer Services influence Hiring



Approach

Explore

- Identified three most critical datasets
- Summarized the data to understand better
- Data imbalance on predictor variable
- Identified sparsely populated variables
- Identified 60 predictor variables

Prepare

- Joined identified datasets
- Derived predictor variables
- Created data dictionary for text data
- Imputed missing values using MICE
- Created train data without imbalance

Predict

- Classification model using Random Forest
- Evaluated the models using AUC
- Cross validation and hyperparameter tuning
- Used Boruta for feature elimination
- Zeroed-in on 32 predictor variables
- Conducted survival analysis based on variable importance

Overview

Dataset	*Contact*	*Hire Info*	*Activities*
# of Categorical variables	264	17	38
# of Continuous variables	0	1	7
# of Textual data variables	42	8	3
# of Unique identifier columns	11	2	12
# of Data / Time variables	74	7	9
Total # of Columns	391	35	69
# of Columns with > 35% 'NA' / Spaces	198	8	6
# of Highly imbalanced columns	17	3	4
# of Variables considered for prediction	36	2	4

Explore

- **Data Filters**
  - **Indicates the account is that of a job seeker**
  - **Active Color Status of client**
    - **Eliminated Black & Blanks**
  - **If the account is active or not**
- **Identified the date and textual columns that can be effectively used**
- **Dropped columns with > 35% missing values**
- **Columns with heavy data imbalance were ignored**
- **Target variable is imbalanced – Hired : 30%, Not Hired: 70%**

http://www.ncsl.org/Portals/1/Documents/Labor/Licensing/vetsFamilies_v03_web.pdf

Prepare

- **Derived target and predictor variables**
  - **Days to Hire a Hired Date – Record Created date**
  - **Resume Complete Duration**
  - **Initial Assessment Complete Duration**
  - **Number of Activities**
- **Data Dictionary - Created data dictionary for fields such as Highest Level of Education, Desired Industry for Employment**
- **Imputation - Columns with missing values < 35% was imputed using MICE with 5 iterations**
- **Under sampling was done on train dataset to make data balanced. Test dataset is left untouched**

Predict

- **Selected Random Forest classification model after comparing different classification models**
- **Hyper parameter tuning and Cross validation**
- **Models were evaluated using AUC**
- **Three variables were removed based on Boruta**
- **Final predictor variables – 35**
- **Survival analysis was conducted on 5 variables that influences model the most**

Analytical Tools

Some of the variables which were found statistically significant using Logistic Regression & Boruta

Variable	*Variable Description*
Gender	Gender of client
Number of Activities	Count of logged activities per client
Highest_Level_of_Education_Completed_c	Education attained
Willing_to_Relocate__c	Willingness to relocation
Resume_Tailoring_Tips__c	Client acquired support to tailor resume
HHUSA_Workshop_Participant__c	Client is participant of HH USA training workshop

Analytical Tools

Algorithm Employed	*Parameter/Hyperparameters*	*Performance Measures*
Random Forest	**NA**	**AUC = 0.65, Accuracy = 57%**
Logistic Regression	**NA**	**AUC = 0.64, Accuracy = 55%**

Data Used	**Target Variable: Hired/NotHired**	**Predictors Count : Around 400**

Analytical Tools

Algorithm Employed	*Parameter/Hyperparameters*	*Performance Measures*
Naive Bayes (Baseline)	**10-fold Cross Validation**	**AUC = 0.8341**
Neural Networks	**size=3, maxit=1000**	**AUC = 0.8891**
GLM Net	**3-fold Cross Validation**	**AUC = 0.9053**
GBM	**3-fold Cross Validation**	**AUC = 0.9056**
Random Forest	**Default Values**	**AUC = 0.9195**

Data Used	**Target Variable: Hired/NotHired**	**Predictors Count : 32**

Analytical Tools

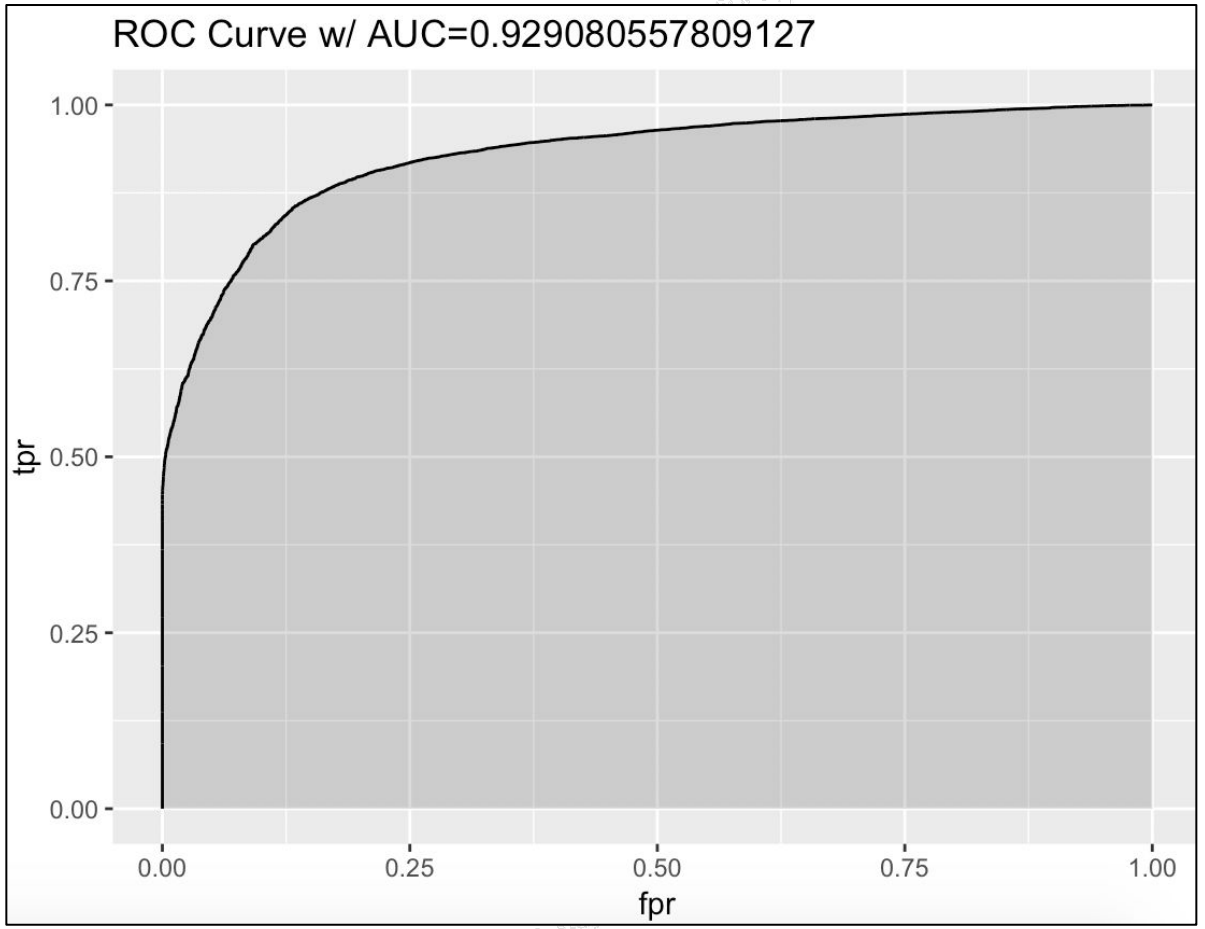
Random Forest Variants (Parameter/Hyperparameters)	*Performance Measures*
Fine-tuning (120 trees, 5 nodes)	**AUC = 0.9182**
10 fold Repeated Cross-Validation, mtry =7, num.tree = 120	**AUC = 0.9211**
10- fold Repeated Cross-Validation (repeats 3 times) splitrule = gini and min.node.size = 3, metric = ROC, mtry =8 (using Tune grid)	**AUC = 0.9263**

Data Used	**Target Variable: Hired/NotHired**	**Predictors Count : 32**

Final AUC & Classification Matrix

	Reference		
Prediction		**Hired**	**Not Hired**
	Hired	**5713**	**1473**
	Not Hired	**1559**	**14136**

Top Features	**Requested_Support__c, Open_duration, Resume_Complete_Duration, Num_Activities, Initial_Assessment_Complete_Duration, Gender__c, Mileage_Willing_To_Commute__c, Used_Volunteer_Services__c, MyTrak_VTS_Assigned__c,**



Results: Important features

Staff_Assigned	*Hired*	*Not Hired*
No	31%	68%
Yes	76%	23%

Reason_for Requested Assistance (top 3)	*%age of total requests*	*Hired*	*Not Hired*
Mock Interviews	around 30%	66%	33%
Industry Specific Job Search	around 26%	72%	27%
LinkedIn Review	around 15%	61%	38%

Results: Important features

- **Classification: Hired = If hired in <180 days, else Not-hired**

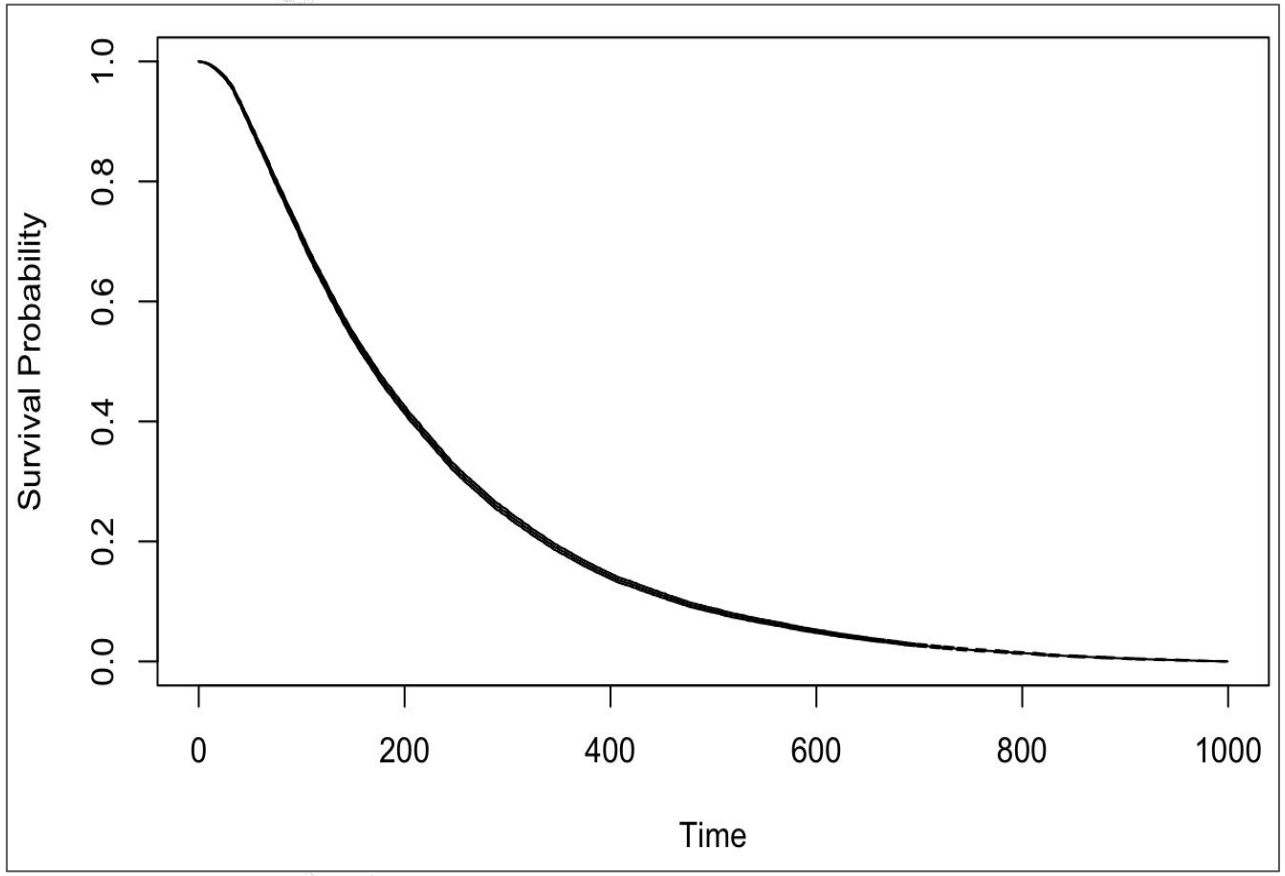
Feature	*Impact*	*Description*
Number of Activities	**Positive**	**The larger the number of activities performed, the greater the chance of being hired**
Transition Specialist Assigned	**Positive**	**If a specialist is assigned, then a greater chance in being hired than when a specialist is not assigned**
Finalized Revised Resume on File	**Negative**	**If a final resume is not on file, then prob of getting hired decreases than when a resume is on file**
Initial Assessment Complete Duration	**Negative**	**The longer the duration in completing the initial assessment, the lower the chances of being hired**
Resume Tailoring Tips	**Negative**	**If client did not receive resume tailoring tips, then a lower chance of being hired than when client did receive resume tailoring tips**
HHUSA Workshop Participant	**Positive**	**If client received attended HHUSA workshop, then a greater chance of being hired than when client did not attend HHUSA workshop**

Results: Demographic features

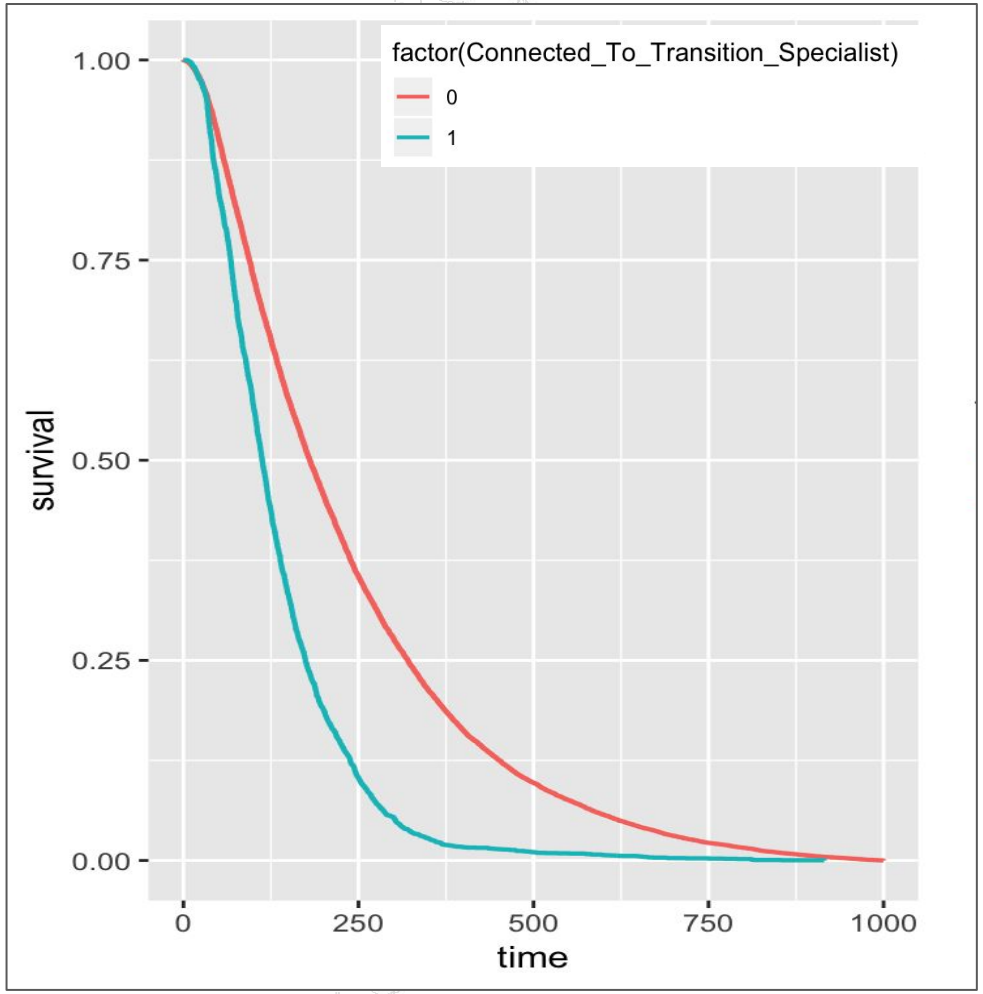
- **Classification: Hired = If hired in <180 days, else Not-hired**

Feature	*Impact*	*Description*
Gender**	**Positive**	**A male client has a higher probability of getting hired than a female client**
Race	**Not Significant**	**Cannot Interpret**
Highest Level of Education**	**Positive**	**A client with a master degree has a higher probability of getting hired than a client without any degree**
Disability > 60%**	**Negative**	**A person with a disability of >60% is less likely to be hired than a person without a disability**

Results: Hire Heroes Impact



Overall Survival Probability



Recommendations

- **Interact more..!**
- **Assign a transition specialist at the earliest**
- **Assist client with resume tailoring tips to have a completed resume on file**
- **Encourage all clients to attend HHUSA workshop**
  - **Mock Interview - top request from client**
- **Host employment bootcamps**

Future work

- **Have a framework to increase the completeness and correctness of data capture**
- **Analyze some of the critical fields such as Rank, Employed Position, Skillset of hired veterans to aid the unemployed veterans**
- **Analyze the funding capabilities and its influence**
- **Personalized career coaching to the client on site (as of now most of the training is virtually offered to clients)**

Thank You!