Master Thesis Defense

Ensemble Machine Learning to Predict Family Consent for Organ Donation

Ehsan Khan

Department of Systems Science and Industrial Engineering

April 20, 2018

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Thomas J. Watson
School of Engineering
and Applied Science

Outline

Introduction

- Increasing demand for organ transplantation
  - Increased incidence of vital organ failure [Cohen et al. (2017)]
  - Substantial disparity between organ demand and supply [OPTN]

114,686
Waiting List as on April 19, 2018

34,770
Organs Transplanted in 2017

20
Die each day from Waiting List

Avg. Wait Time

5 years
Kidney

2 years
Pancreas

11 months
Liver

4 months
Heart & Lung

- Measures taken to bridge the gap:
  - Aggressive organ utilization
  - **Increasing family consent for Organ Donation (OD)**

Research Motivation

- Early Interaction (EI) project
  - Valuable comments during quarterly review meeting
  - On-site visit to Lincoln Hospital
  - Staff cooperation
  - Team effort in data validation and correction
- No literature on Early Support to family
- Limited literature emphasize the application of modeling techniques to predict family consent

Research Objectives

- Build a family **consent prediction model**
  - Evaluate effect of EI on CR
  - Identify significant factors of CR
- Develop real-time consent prediction before approach
- Propose staff dispatch algorithm
  - Identify top 3 approachers for each pending case

Research Questions

- Does EI with family have impact on CR?
- What are significant factors of CR for 2016-2018 **LiveOn**NY data?
- How accurately can we predict consent before approach?
- Are there any evidences if a particular approacher is more likely to get consent over others?

Research Concepts and Contributions

Studied Concepts

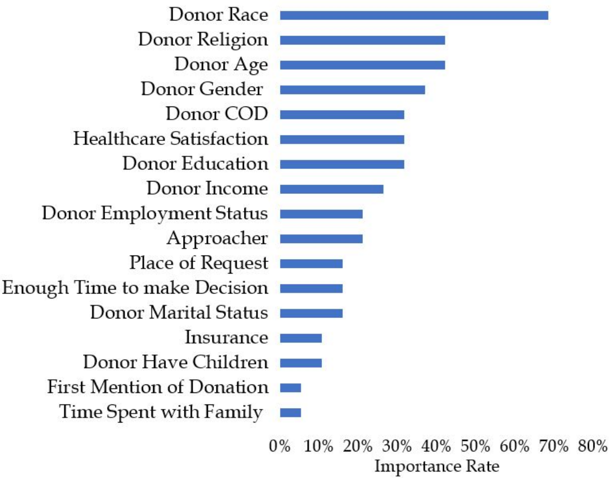
- CR Factors in Literature
- Feature Transformation
- Prediction Modeling Methods
- Model Deployment to Production

Research Contribution

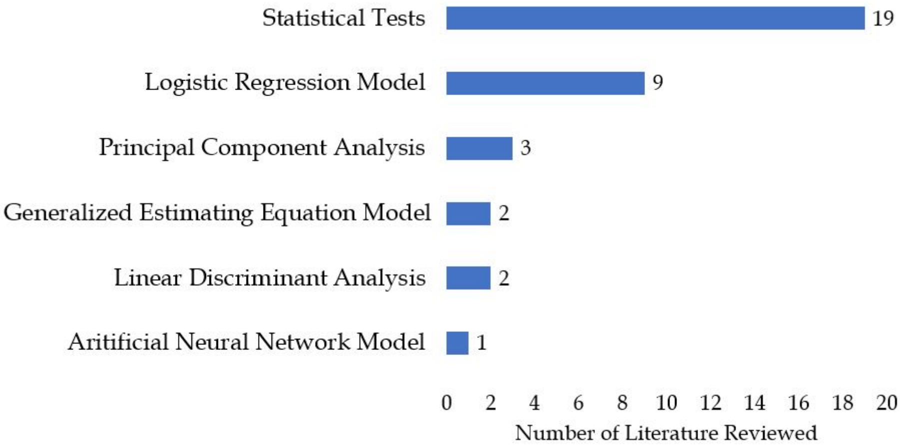
- Early Support Impact
- Advance Consent Prediction

CR Factors - Literature Review

- 19 literature reviewed in U.S. setting



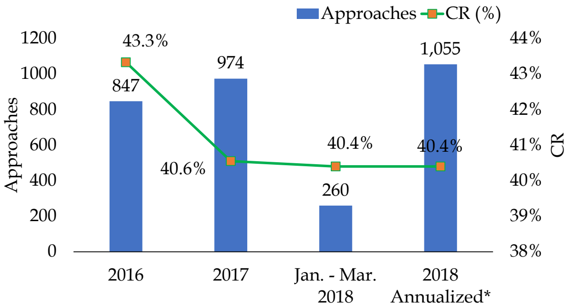
Methodology Literature Review



Data Summary (Jan. 2016 - Mar. 2018)

	Total	2016	2017	Jan. - Mar. 2018	2018 Annualized*
Referral	**9,694**	**4,011**	**4,354**	**1,329**	**5,390**
Approaches	2,081	847	974	260	1,055
Donor	653	271	296	86	349
Consent Not Recovered (CNR)	214	96	99	19	77
No Authorization	1,214	480	579	155	629

*As on 04/19/2018



Data Description

- **LiveOn**NY data for all approaches (Jan. 2016 - Mar. 2018)
- Sample size: N = 2,079 (Approaches)
- 29 factors grouped under
  - Donor (6)
  - Next-of-kin (2)
  - Hospital (4)
  - OPO related (17)
- Family consent to organ donation (Yes/No)

CR Factors for Prediction Model

Deceased donor factors:

- Age, gender
- Race, religion
- FPA, COD

Next-of-kin factors:

- Gender, relationship

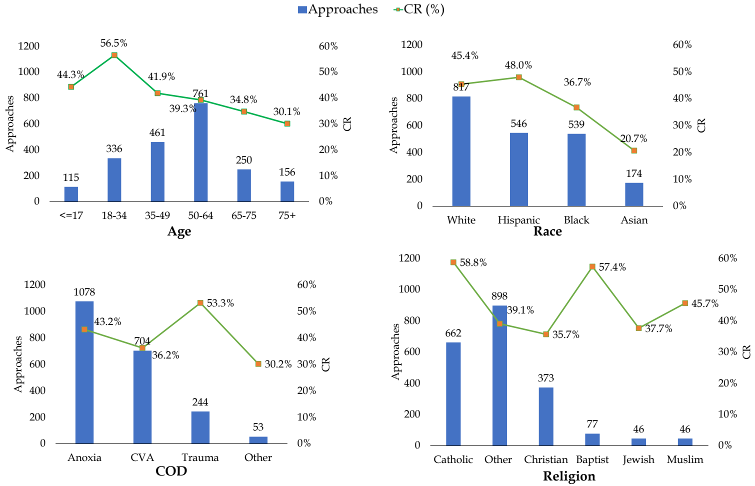
Hospital factors:

- Unit, county, system
- Timely referral

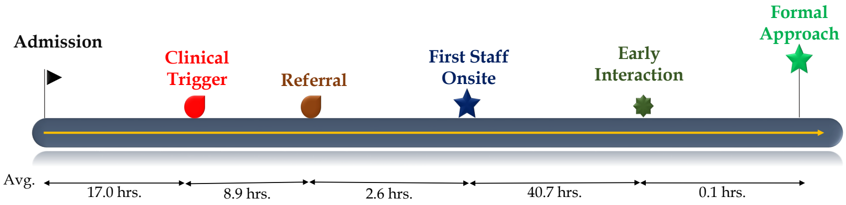
OPO factors:

- Number of conversation with family by OPO staff before formal approach
- Number of staff involved per case before formal approach
- Initial mention to family about OD
- Title, team, CTB, involve hospital
- Discuss benefits, private setting
- Referral to first staff on-site & formal approach
- Referral to BD declaration
- BD declaration to formal approach
- Donation mention prior to death, arrival, referral, OPO speak with family

Donor Factors Summary



Admission to Formal Approach Timeline



Timing	Avg. Hours	
	Consented	Declined
Admission to Clinical Trigger	15.5	18.4
Clinical Trigger to Referral	8.0	9.8
Referral to First Staff Onsite	2.6	2.7
First Staff Onsite to Early Interaction	41.6	39.8
Early Interaction to Formal Approach	0.1	0.1

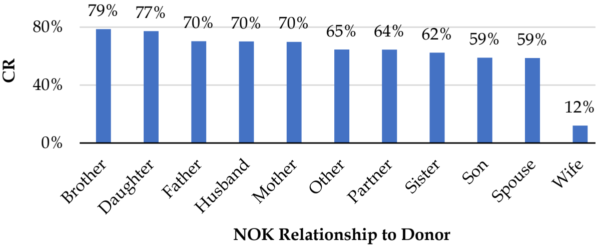
Timing Analysis

		Approaches	CR (%)	P-Value
Admission to Referral	0-24 hrs.	886	47.7	<0.05
	>24 hrs.	1,149	36.8	
Clinical Trigger to Referral	<0 hrs.	351	51.3	<0.05
	0-1 hrs.	170	45.3	
	1.1-7 hrs.	419	53.7	
	>7 hrs.	759	44.8	
Formal Approach	Before BD (22.9 hrs.)	245	82.0	<0.05
	After BD (10.9 hrs.)	758	51.7	
Referral to First Staff Onsite	0-4 hrs.	1,457	44.3	<0.05
	>4 hrs.	289	35.9	
First Staff Onsite to Formal Approach	0-24 hrs.	621	41.6	<0.05
	>24 hrs.	1,041	46.8	
# Conversation before Formal Approach	No Conversation	456	37.5	<0.05
	1	590	44.9	
	>1	355	59.3	

Factor labels with higher CR

	Donor	CNR	No Authorization
Average # Conversation before Formal Approach	2	1	1
Average # Staff Involved per case before Formal Approach	5	4	4

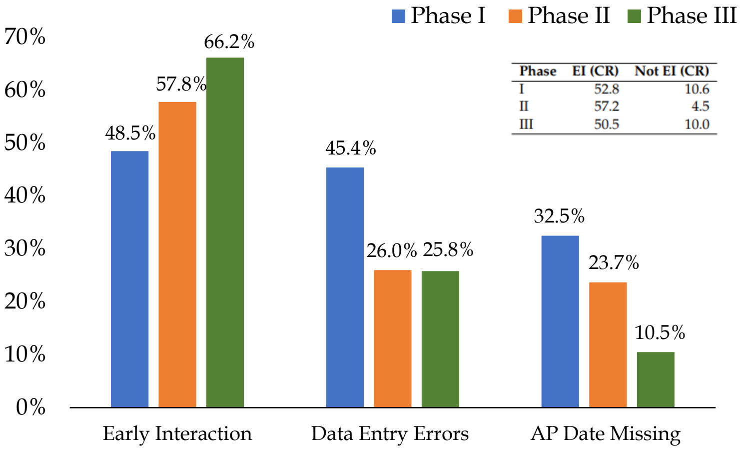
Factor	Level	CR(%)
Donor Gender	Male	43.4
	Female	39.3
FPA	Yes	85.7
	No	36.0
NOK Gender	Male	68.5
	Female	66.8



Early Interaction

- Early Interaction
  - At least one conversation with family before formal approach
- Data Entry Errors
  - Initial Contact with AP date after Formal Approach date

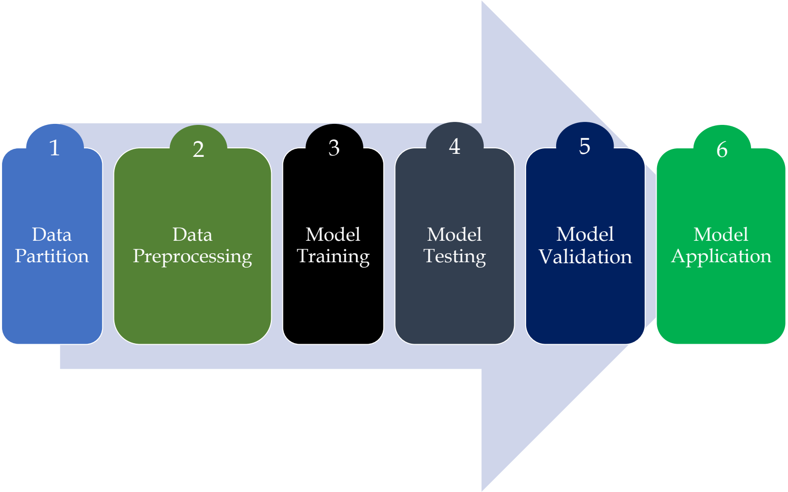
Early Interaction Progress



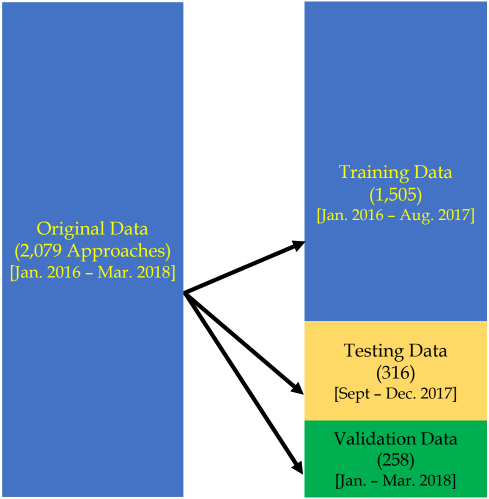
Phase	EI (CR)	Not EI (CR)
I	52.8	10.6
II	57.2	4.5
III	50.5	10.0

- **Work In Progress:** Data Collection for EI PDSA

6-Step Model Building Process



Step 1: Data Partition



Step 2: Data Preprocessing

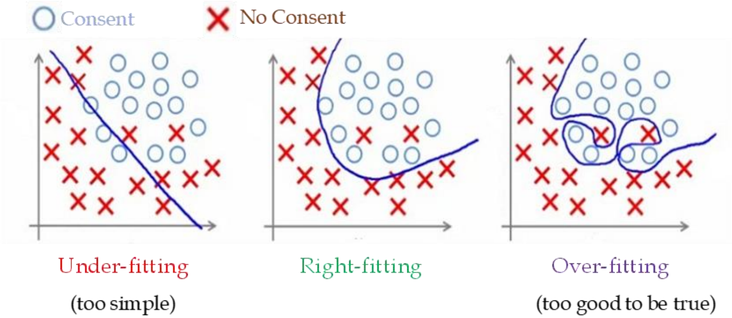
- 5 factors removed (such as donation mentioned prior to death)
  - 50% missing values
- Factor transformation (Example: Hospital Team)

	Label Encoding	Normalization
C	1	0.2
L	2	0.4
N	3	0.5

- **Linear Discriminant Analysis (LDA)**: compress information retained in all factors

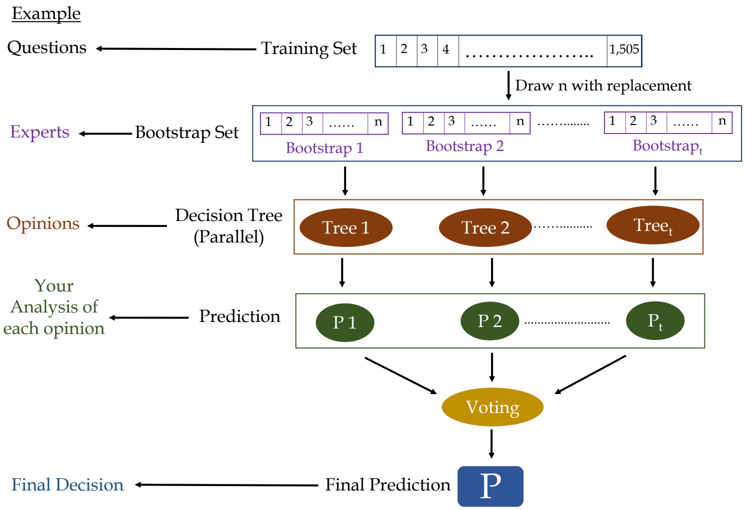
Step 3: Model Training

- Ensemble Machine Learning:
  - Combine many models to improve prediction power
- How does the machine learn itself?

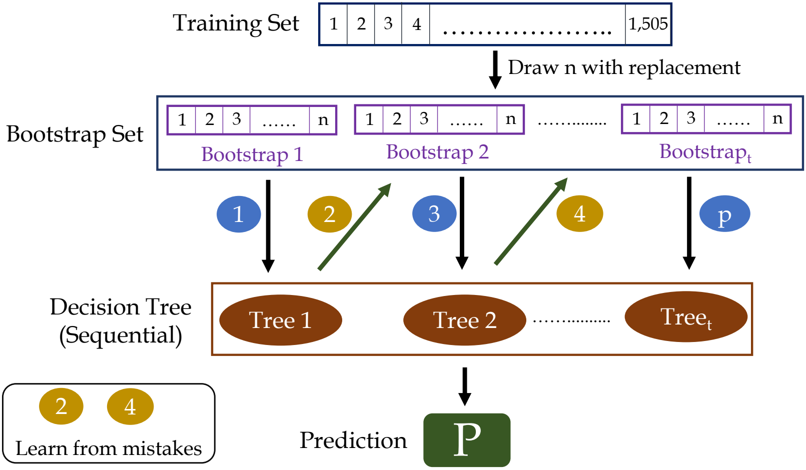


- **Bagging & Boosting**

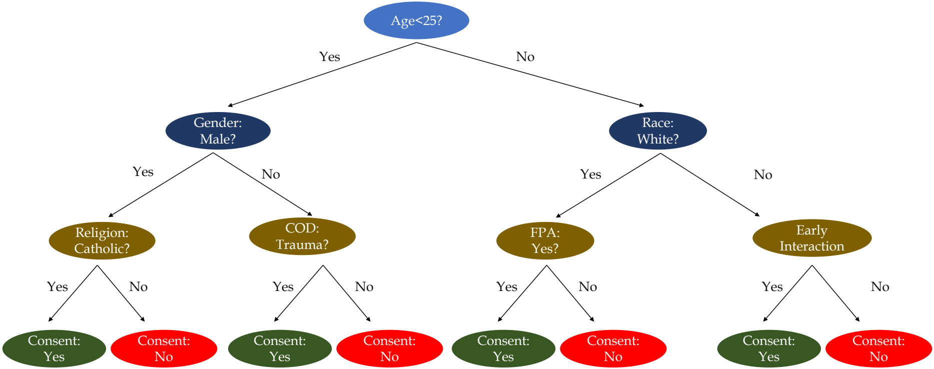
Bagging Technique (Bootstrap Aggregation)



Boosting Technique



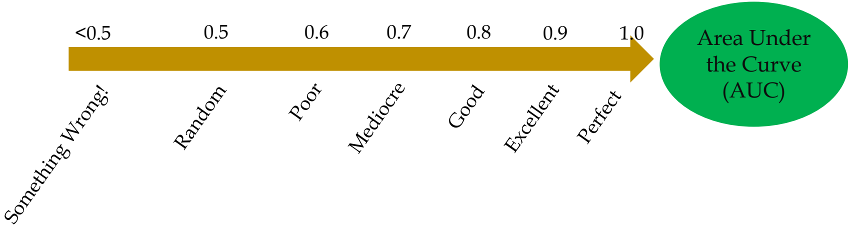
Building Decision Tree



Machine Learning Models

- Bagging **(Parallel - Opinion Model)**
  - Random Forest (Ho, 1995)
  - Extra Tree (Geurts et al., 2006)
- Boosting **(Sequential - Opinion Model)**
  - AdaBoost (Freund and Schapire, 1995)
    - Weak learners combined into a weighted sum
  - Gradient Boosting (Breiman, 1997)
    - Weak learn performance is improved by iteratively shifting focus towards incorrect prediction
  - eXtreme Gradient Boosting (Chen, 2016)
    - Regularized objective function for better model performance

Step 4: Model Testing



- Model results with test data (N = 316, Sept. - Dec. 2017)

Table 5.1: Final Prediction Test AUC

Ensem. Techniq.	Model	Without Preprocessing	Normalization + LDA	Parameter Tuning
Boosting	**XGB**	0.8197	0.8415	**0.8946**
	GB	0.8228	0.8348	0.8705
	AB	0.8078	0.7985	0.8038
Bagging	ETree	0.8124	0.8029	0.8465
	RF	0.7458	0.8403	0.8325

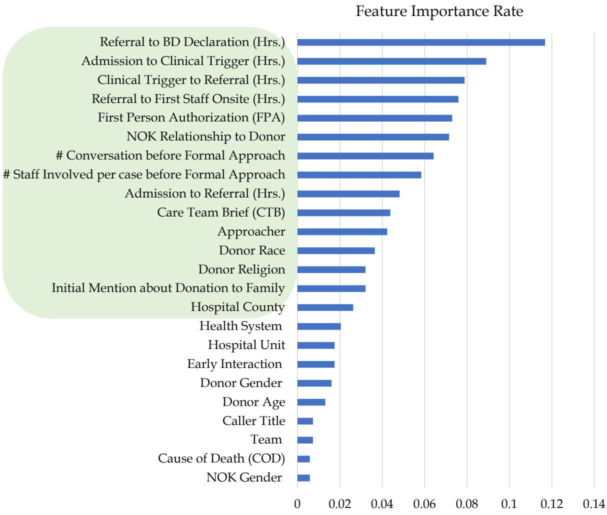
Step 5: Model Validation

- Training: Jan. 2016 - Aug. 2017
- Prediction: Jan. - Mar. 2018

Case	Actual Outcome (Consent)	Predicted Outcome from Model (Consent)	Probability of Consent
1	Yes	**Yes**	0.7824
2	No	**Yes**	0.5236
3	Yes	**Yes**	0.7028
4	Yes	**Yes**	0.7361
5	Yes	**Yes**	0.8830
6	No	**Yes**	0.5825
7	Yes	**Yes**	0.6089
8	Yes	**Yes**	0.7702
9	Yes	**Yes**	0.8009
10	Yes	**Yes**	0.9008
....	....	....	....
....	....	....	....
258	Yes	**No**	0.4125

- 223/258 cases correctly predicted (Accuracy: **92.8%!!!**)

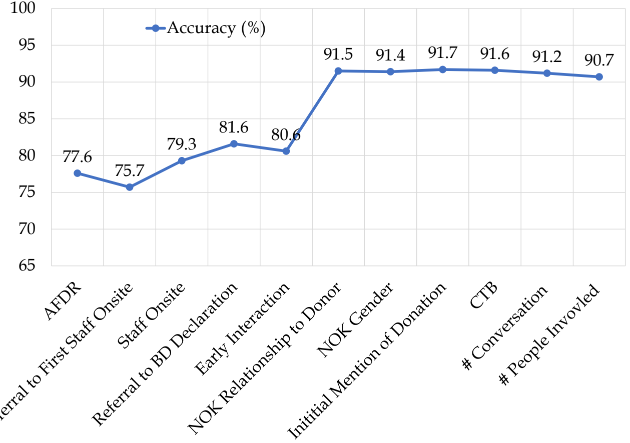
Important Factors by XGB Model



Factors at Referral and before Formal Approach

Available Factors at Referral (AFDR)	**As Information entered in iTx**
Donor Age	Referral to First Staff Onsite (hrs.)
Donor Gender	Staff who is Onsite
Donor Race	Referral to BD Declaration (hrs.)
Donor Religion	Early Approach
FPA	NOK Relationship to Donor
COD	NOK Gender
Hospital Team	Initial Mention of Donation to Family
Hospital County	CTB
Health System	# Conversation before Formal Approach
Admission to Clinical Trigger (hrs.)	
Clinical Trigger to Referral (hrs.)	

Model Performance as Factors are added after Referral



Step 6: Model Application

- Calculated probability of family consent if different staff approach the same case

Case	**Actual Approacher**	**A1**	A2	**A3**	A4	**A5**	**Model Recommendation**
1	0.6772	0.7772	0.6772	0.5772	0.7672	**0.7792**	A5
2	0.7761	0.7561	**0.7861**	0.7161	0.7261	0.7061	A2
3	0.6268	0.5737	0.5875	0.5146	**0.7046**	0.6768	A4
4	0.8780	0.6773	0.7780	**0.8880**	0.7278	0.7599	A3
5	0.1053	0.3022	0.4053	**0.4853**	0.2105	0.3505	A3

Model Recommendation

- 10 pending cases on April 16, 2018

Table 5.2: Probability of Family Consent

Case	A1	A2	A3	A4	A5	Who should Approach Family?
1	0.2086	0.2768	**0.2906**	0.2551	0.2551	A3
2	0.3682	**0.5394**	**0.5394**	0.4262	0.4262	A2, A3
3	0.2263	**0.3703**	**0.3703**	0.3188	0.3188	A2, A3
4	0.1273	**0.3577**	**0.3577**	0.2951	0.2951	A2, A3
5	**0.6250**	0.6230	0.6230	0.5680	0.5680	A1
6	0.3845	**0.6807**	**0.6807**	0.6158	0.6158	A2, A3
7	0.2292	0.3269	**0.3428**	0.2816	0.2816	A3
8	0.5254	0.6429	0.6429	**0.6479**	**0.6479**	A4,A5
9	0.3403	**0.5124**	**0.5124**	0.4413	0.4413	A2,A3
10	0.2027	**0.3712**	**0.3712**	0.3196	0.3196	A2,A3

A1, A2, A3, A4, and A5 are FSCs who are on schedule for next three days

Conclusion

- Highlighted Early Interaction importance
- Boosting (Sequential - Opinion Model) gave best result
- Application of dynamic modeling approach
- Significant factors are identified using XGB model
- Recommendation from approacher dispatch model

Future Work

- Make dispatch model more efficient
- Fully automated consent prediction dashboard
- Complete documentation of important variables in iTx

Acknowledgements

- Dr. Sang Won Yoon (Advisor)
- Dr. Mohammad Khasawneh (Director of Graduate Studies of ISE)
- **LiveOn**NY
  - Helen Irving (President and CEO)
  - Amy L. Friedman M.D. (CMO and Exec. VP)
  - Kristin Delli Carpini (Director, Performance Improvement)
  - Emmanuel Marquez (Clinical Data Analyst) - Mentor
  - Pradeep Bagavan (Systems Data Analyst) - Mentor
  - Natalie Benavides (VP, Ext. Affairs and Strat.Dev., Hosp. Services)
  - Lizbeth Waddell (Director, Hospital and Family Services)
  - Walesca Marmolejos (Assistant Director, Family Services)
  - Jeff Lewis (VP, Clinical Operations)
  - Kervens Louissaint (Director, Clinical Services)
  - Richard Budke (Asisstant Director, Clinical Services)
  - Jennifer Dillon (Asisstant Director, Hospital Services)

Questions?

References I

Chandler, J. A., M. Connors, G. Holland, and S. D. Shemie (2017). âĂIJeffectiveâĂİ requesting: A scoping review of the literature on asking families to consent to organ and tissue donation. *Transplantation 101*(5S), S1–S16.

Ghorbani, F., H. Khoddami-Vishteh, O. Ghobadi, S. Shafaghi, A. R. Louyeh, and K. Najafizadeh (2011). Causes of family refusal for organ donation. In *Transplantation Proceedings*, Volume 43, pp. 405–406. Elsevier.

Godin, G., A. Bélanger-Gravel, C. Gagné, and D. Blondeau (2008). Factors predictive of signed consent for posthumous organ donation. *Progress in Transplantation 18*(2), 109–117.

References II

Goldberg, D. S., S. D. Halpern, and P. P. Reese (2013). Deceased organ donation consent rates among racial and ethnic minorities and older potential donors. *Critical Care Medicine 41*(2), 496.

Hong, J. C., H. Yersiz, P. Kositamongkol, V. W. Xia, F. M. Kaldas, H. Petrowsky, D. G. Farmer, G. Lipshutz, D. Markovic, J. R. Hiatt, et al. (2011). Liver transplantation using organ donation after cardiac death: A clinical predictive index for graft failure–free survival. *Archives of surgery 146*(9), 1017–1023.

Irving, M. J., A. Tong, S. Jan, A. Cass, J. Rose, S. Chadban, R. D. Allen, J. C. Craig, G. Wong, and K. Howard (2011). Factors that influence the decision to be an organ donor: a systematic review of the qualitative literature. *Nephrology Dialysis Transplantation 27*(6), 2526–2533.

References III

Mostafa, M. (2008). Profiling organ donors in egypt using intelligent modeling techniques. *Marketing Intelligence & Planning 26*(2), 166–188.

Rabinstein, A. A., A. H. Yee, J. Mandrekar, J. E. Fugate, Y. J. de Groot, E. J. Kompanje, L. A. Shutter, W. D. Freeman, M. A. Rubin, and E. F. Wijdicks (2012). Prediction of potential for organ donation after cardiac death in patients in neurocritical state: a prospective observational study. *The Lancet Neurology 11*(5), 414–419.

Robbins, M. L., D. A. Levesque, C. A. Redding, J. L. Johnson, J. O. Prochaska, M. S. Rohr, and T. G. Peters (2001). Assessing family membersâĂŹ motivational readiness and decision making for consenting to cadaveric organ donation. *Journal of Health Psychology 6*(5), 523–535.

References IV

Rodrigue, J. R., D. L. Cornell, and R. J. Howard (2006). Organ donation decision: comparison of donor and nondonor families. *American Journal of Transplantation 6*(1), 190–198.

Schleich, B. R., S. S. Lam, S. W. Yoon, W. Tajik, M. J. Goldstein, and H. Irving (2013). A neural network-based approach for predicting organ donation potential. In *IIE Annual Conference. Proceedings*, pp. 1532. Institute of Industrial and Systems Engineers (IISE).

Shah, M. B., V. Vilchez, A. Goble, M. F. Daily, J. C. Berger, R. Gedaly, and D. A. DuBay (2018). Socioeconomic factors as predictors of organ donation. *Journal of Surgical Research 221*, 88–94.

Siminoff, L. A., N. Gordon, J. Hewlett, and R. M. Arnold (2001). Factors influencing families' consent for donation of solid organs for transplantation. *Jama 286*(1), 71–77.

References V

Simpkin, A. L., L. C. Robertson, V. S. Barber, and J. D. Young (2009). Modifiable factors influencing relativesâĂŹ decision to offer organ donation: systematic review. *Bmj 338*, b991.

Walker, W., A. Broderick, and M. Sque (2013). Factors influencing bereaved familiesâĂŹ decisions about organ donation: an integrative literature review. *Western Journal of Nursing Research 35*(10), 1339–1359.

Webb, G., N. Phillips, S. Reddiford, and J. Neuberger (2015). Factors affecting the decision to grant consent for organ donation: a survey of adults in england. *Transplantation 99*(7), 1396–1402.

Publication List

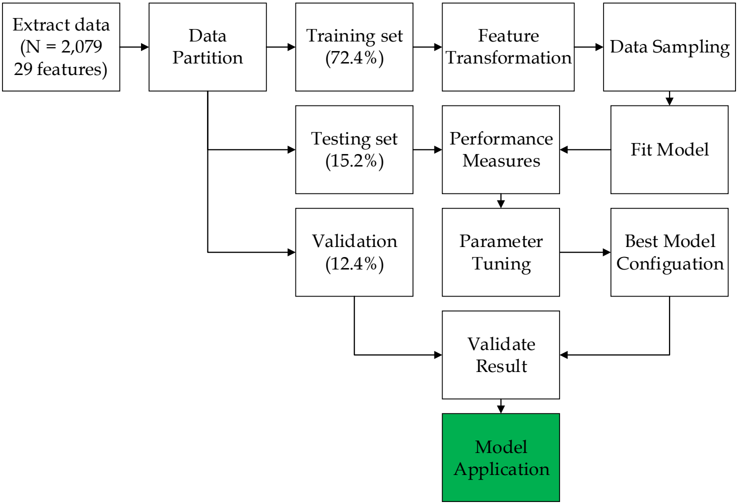
Khan, M.E., Choudhury, Friedman, A., Won, D. (2018) "Decision Support System for Renal Transplantation." IIE Annual Conference. Proceedings. Institute of Industrial and Systems Engineers (IISE), 2018

Halawa, F., Lee, I.G., Shen, W., Khan, M.E. (2017) "The Implementation of Hybrid MTS/MTO as a Promoter to Lean-Agile: A Simulation Case Study for Miba Sinter Slovakia." IIE Annual Conference. Proceedings. Institute of Industrial and Systems Engineers (IISE), 2017

Literature Review

Year	Publications
2018	Shah et al. (2018)
2017	Chandler et al. (2017)
2015	Webb et al. (2015)
2014	DuBay et al. (2014)
2013	Goldberg et al. (2013), Schleich et al. (2013), Walker et al. (2013)
2012	Andrews et al. (2012), Quick et al. (2012), Rabinstein et al. (2012), Resnicow et al. (2012)
2011	Ghorbani et al. (2011), Irving et al. (2011)
2009	Simpkin et al. (2009)
2008	Mostafa (2008)
2006	Rodrigue et al. (2006)
2002	Minniefield et al. (2002)
2001	Robbins et al. (2001), Siminoff et al. (2001)

Model Building Process Flow



CR Factors from Literature (1 of 2)

Factors	Publications
Age	Rodrigue et al. (2006), Padela et al.(2011), Goldberg et al. (2013), Webb et al. (2015)
Gender	Rodrigue et al. (2006), Webb et al. (2015)
Ethnicity	Rodrigue et al. (2006), Padela et al. (2011), Goldberg et al. (2013), Webb et al. (2015), Chandler et al. (2017)
Religion	Ghorbani et al. (2011), Irving et al. (2011), Webb et al. (2015), Chandler et al. (2017)
Marital Status	Rodrigue et al. (2006), Webb et al. (2015)
Education	Rodrigue et al. (2006)
Have Children	Webb et al. (2015)
Amount of Time Spent	Padela et al. (2011)
Place of Request	Rodrigue et al. (2006), Simpkin et al. (2009)

CR Factors from Literature (2 of 2)

Factors	**Publications**
Employment Status	Rodrigue et al. (2006)
Insurance	Padela et al. (2011)
Cause of Death	Rodrigue et al. (2006), Padela et al. (2011)
First Mention of Donation	Rodrigue et al. (2006)
Donation Requester	Rodrigue et al. (2006), Simpkin et al. (2009), Padela et al. (2011), Chandler et al. (2017)
Timing of Initial Donation Discussion	Simpkin et al. (2009), Chandler et al. (2017)
Overall Satisifaction of Healthcare Team	Rodrigue et al. (2006), Simpkin et al. (2009), Irving et al. (2011), Chandler et al. (2017)
Given Enough Time to make decision	Simpkin et al. (2009), Rodrigue et al. (2006)

Methodologies in Literature

Methods	**Publications**
Statistical Tests	Simpkin et al. (2009), Hong et al. (2011), Goldberg et al. (2013), Webb et al. (2015), Chandler et al. (2017)
Logistic Regression (LR)	Siminoff et al. (2001), Rodrigue et al. (2006), Ghorbani et al. (2011), Irving et al. (2011), Godin et al. (2008), Walker et al. (2013), Rabinstein et al. (2012), Webb et al. (2015), Shah et al. (2018)
Principal Component Analysis (PCA)	Robbins et al. (2001)
Linear Discriminant Analysis (LDA)	Mostafa (2008)
Artificial Neural Network (ANN)	Mostafa (2008), Schleich et al. (2013)

Automated Machine Learning Tool: TPOT

- Tree-Based Pipeline Optimization (TPOT), uses genetic algorithms
- Test AUC after 100 generations: 0.8789

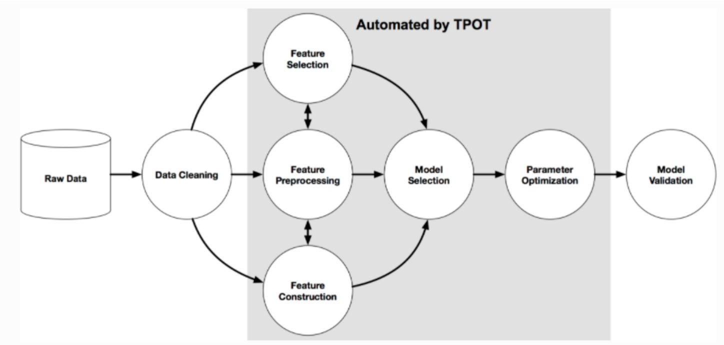


Figure 5.1: An example of machine learning pipeline