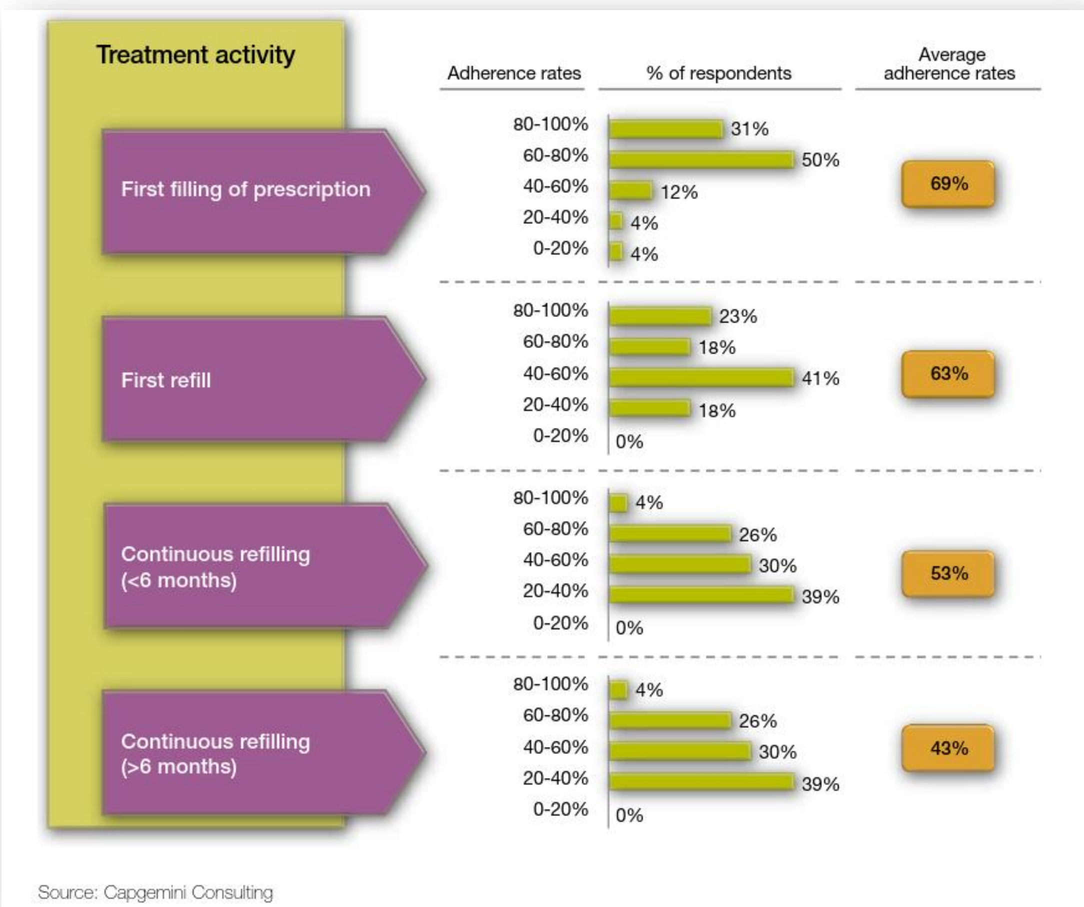
Project Report

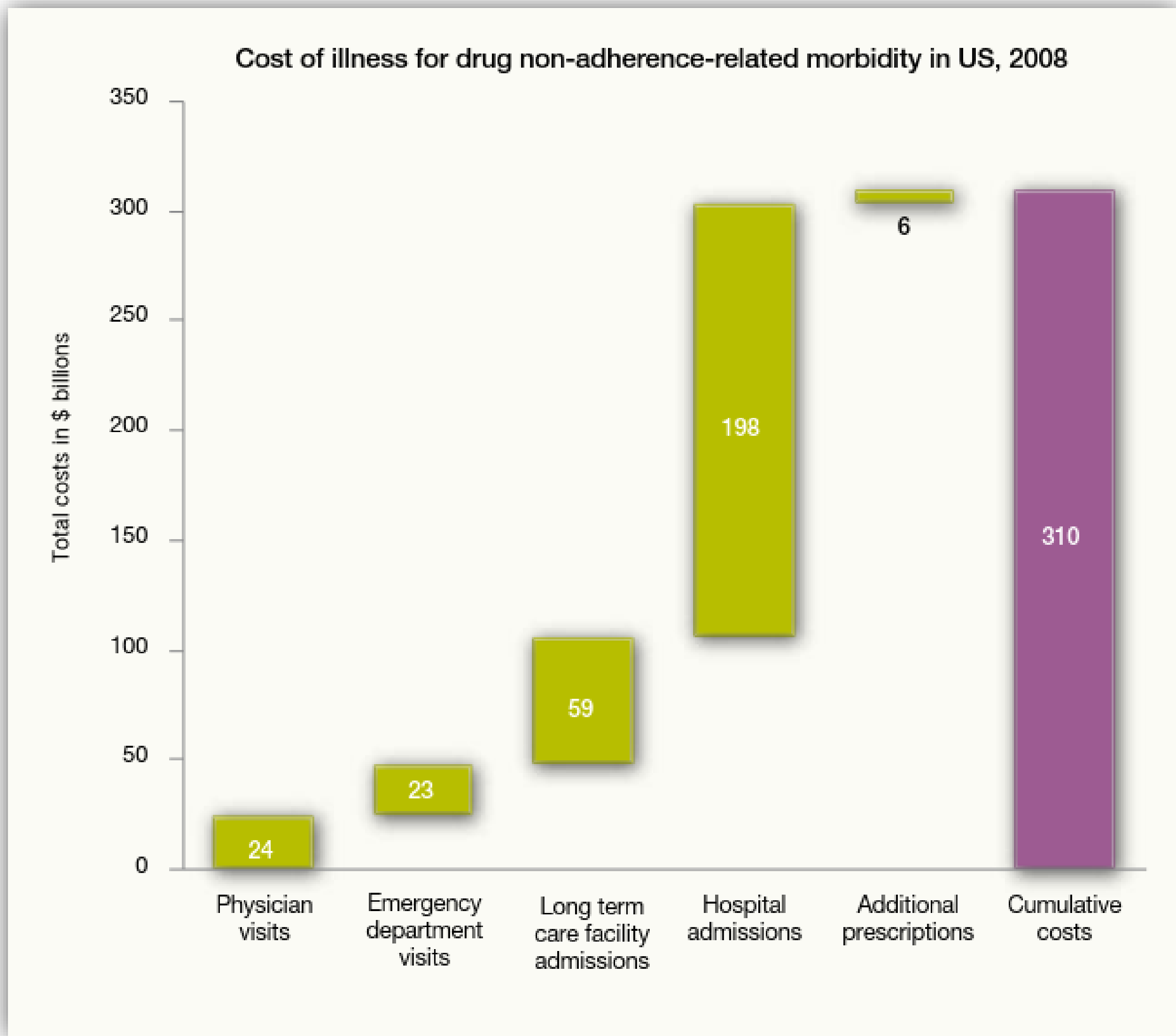
Developing Robust Patient Adherence Model

Shivam Kanoria

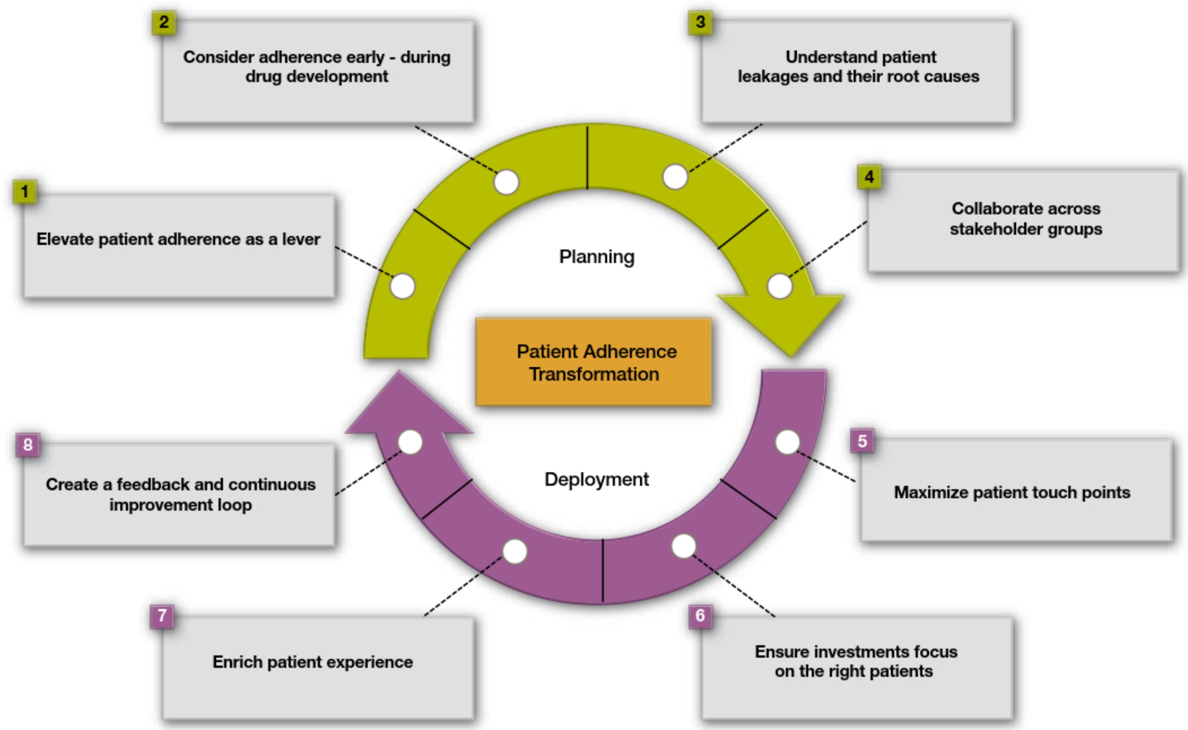
MEDICAL ADHERENCE

- Next frontier in healthcare innovation
- Medical adherence is the most vital element in bridging the gap between the patient’s health and the evolution in medical sciences.
- Platforms for exchanging info increases the ability of the healthcare system to educate patients and importance of adherence
- Intelligent diagnostics linked with monitoring facilities
- Combination of compliance and persistence.
- Compliance is the degree to which a patient follows or completes a prescribed diagnostic




Cost of illness for drug non-adherence-related morbidity in US, 2008
Total costs in $ billions
350
300
250
200
150
100
50
0
24
23
59
198
6
310
Physician visits
Emergency department visits
Long term care facility admissions
Hospital admissions
Additional prescriptions
Cumulative costs

Framework On Patient Adherence



Source: Capgemini Consulting

APPROACH

- These studies can only be transformed into systems when there is a system which can segregates adherent and non-adherents patients. (Classification)
- The client needs a model that can account for the risk of non-adherence per patient and predict if the patient is likely to non-adhere.
- Irregularity in buying the next dosages at the supposed intervals.
- Predicting the **non-adherents** correctly would be the prime objective of the model, so that appropriate actions can be taken by the client.
- Key parameters which would drive the model can be age, income, awareness scale, education, location, family size, severity of ailment etc.
- Several modelling techniques would be needed to check from which the data can be most accurately predicted.

RAW DATASET

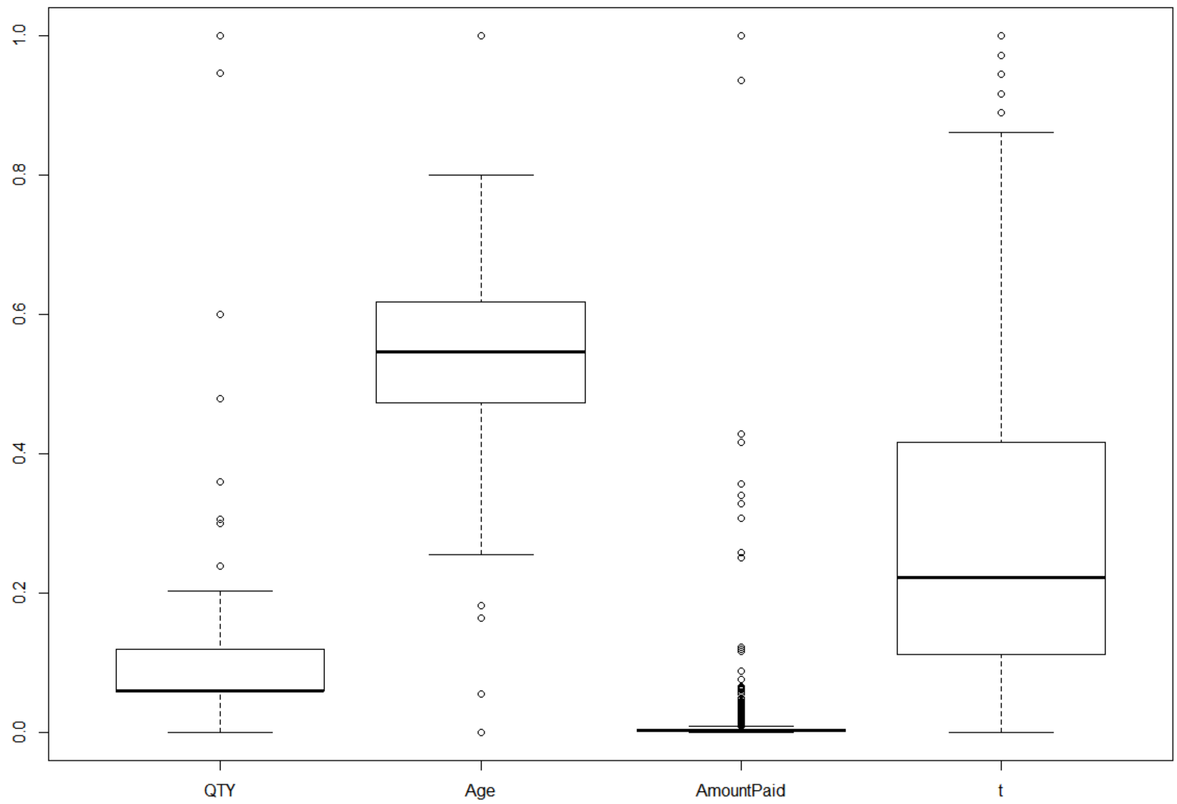
	PatientID	Medication	ActGPI	RouteOfAdmin	State	QTY	For_How_Many_Days	Age	MailRetail	Date	AmountPaid	Sex	PurchasedBy	Pharmacy
1	1001	Medication1	340	A	IN	30.0	30	59	R	30-01-12	7	M	Patient	PHARMACY 1
2	1001	Medication1	340	A	IN	30.0	30	59	R	20-03-12	7	M	Patient	PHARMACY 1
3	1001	Medication1	340	A	IN	30.0	30	59	R	11-05-12	7	M	Patient	PHARMACY 1
4	1001	Medication1	340	A	IN	30.0	30	59	R	24-06-12	7	M	Patient	PHARMACY 1
5	1001	Medication1	340	A	IN	30.0	30	59	R	31-07-12	7	M	Patient	PHARMACY 1
6	1001	Medication1	340	A	IN	30.0	30	59	R	07-09-12	7	M	Patient	PHARMACY 1
7	1001	Medication1	340	A	IN	30.0	30	59	R	11-10-12	7	M	Patient	PHARMACY 1
8	1001	Medication1	340	A	IN	30.0	30	59	R	10-11-12	7	M	Patient	PHARMACY 1
9	1001	Medication1	340	A	IN	30.0	30	59	R	11-01-13	7	M	Patient	PHARMACY 1
10	1001	Medication1	340	A	IN	30.0	30	59	R	09-02-13	7	M	Patient	PHARMACY 1
11	1001	Medication1	340	A	IN	30.0	30	59	R	11-04-13	7	M	Patient	PHARMACY 1
12	1001	Medication1	340	A	IN	30.0	30	59	R	08-05-13	7	M	Patient	PHARMACY 1
13	1001	Medication1	340	A	IN	30.0	30	59	R	09-07-13	7	M	Patient	PHARMACY 1
14	1001	Medication1	340	A	IN	30.0	30	59	R	07-09-13	7	M	Patient	PHARMACY 1
15	1001	Medication1	340	A	IN	30.0	30	59	R	08-10-13	7	M	Patient	PHARMACY 1

- Missing Geographic details
- Transactional records from pharmacies
- Derive features from purchase date

DATASET

Attributes		
Categorical		Numeric
	Factor levels	
PatientID	59	QTY
MedicationID	470	For_How_Many_Days
ActGPI	120	Age
RouteOfAdmin	15	AmountPaid
State	2	
MailRetail	2	
Sex	2	
PurchasedBy	2	
Pharmacy	44	

BOXPLOT



DATA PREPROCESSING – FEATURE ENGG.

- Ordering dataset in chronological order of transaction dates for patient-medication combination.
- Checking for missing values.
- Converting the attributes in factors, numeric and date.
- Removing the redundant level in PurchasedBy attribute.
- Calculating delay for every transaction using Date and For_How_many_days.
- Adding new feature t for giving order of transactions for all patient-medication combinations.
- Removing all rows where number of instances for patient-medication is less than three. (why ?)
- Create dummies from categorical attributes and standardize numeric data.
- Removing Date and For_How_many_days from dataset.
- Removing PatientID, MedicationID, PatientID-MedicationID.

NEW FEATURE – delay, t, target

PatientID	Medication	For_How_Many_Days	Date	delay	t	target
1001	Medication1	30	30-Jan-12	0	1	1
1001	Medication1	30	20-Mar-12	20	2	0
1001	Medication1	30	11-May-12	22	3	0
1001	Medication1	30	24-Jun-12	14	4	0
1001	Medication1	30	31-Jul-12	7	5	0
1001	Medication1	30	07-Sep-12	8	6	0
1001	Medication1	30	11-Oct-12	4	7	0
1001	Medication1	30	10-Nov-12	0	8	1
1001	Medication1	30	11-Jan-13	32	9	0
1001	Medication1	30	09-Feb-13	-1	10	1
1001	Medication1	30	11-Apr-13	31	11	0
1001	Medication1	30	08-May-13	-3	12	1
1001	Medication1	30	09-Jul-13	32	13	0
1001	Medication1	30	07-Sep-13	30	14	0
1001	Medication1	30	08-Oct-13	1	15	1

CONFUSION MATRIX		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

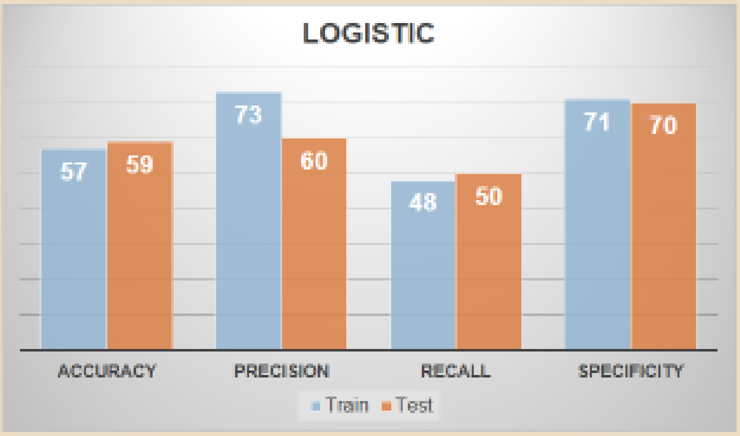
0 = Non-adherent
1 = Adherent

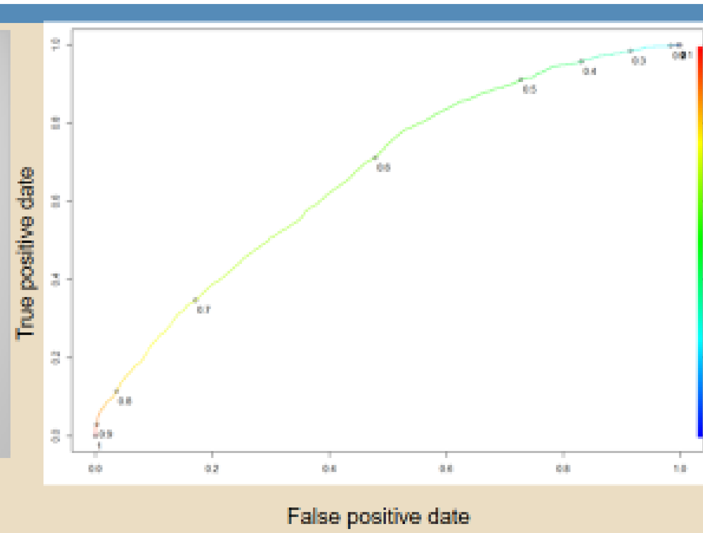
$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+TP+FN+FP}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

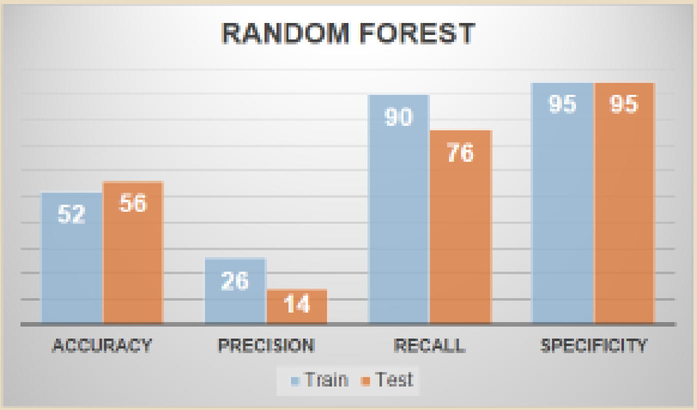
$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

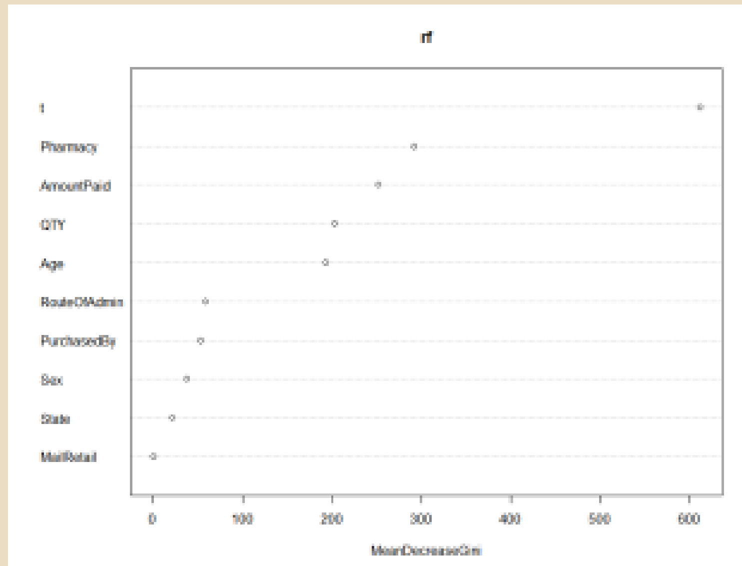
$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}$$

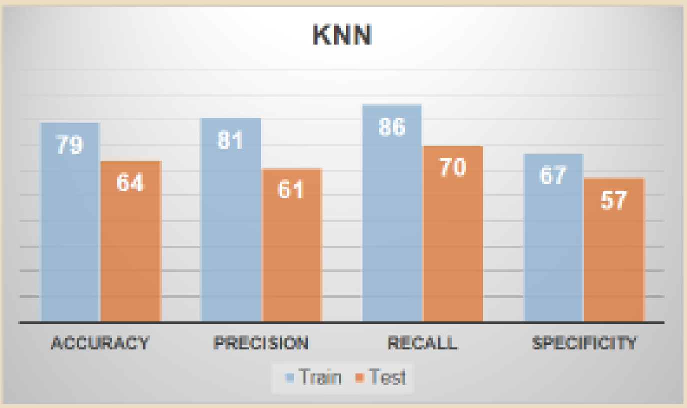


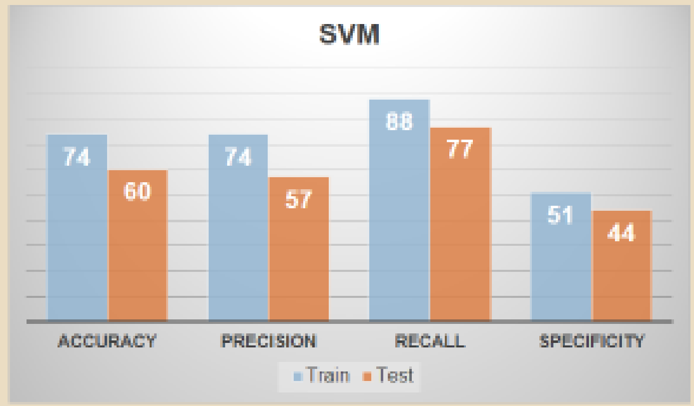


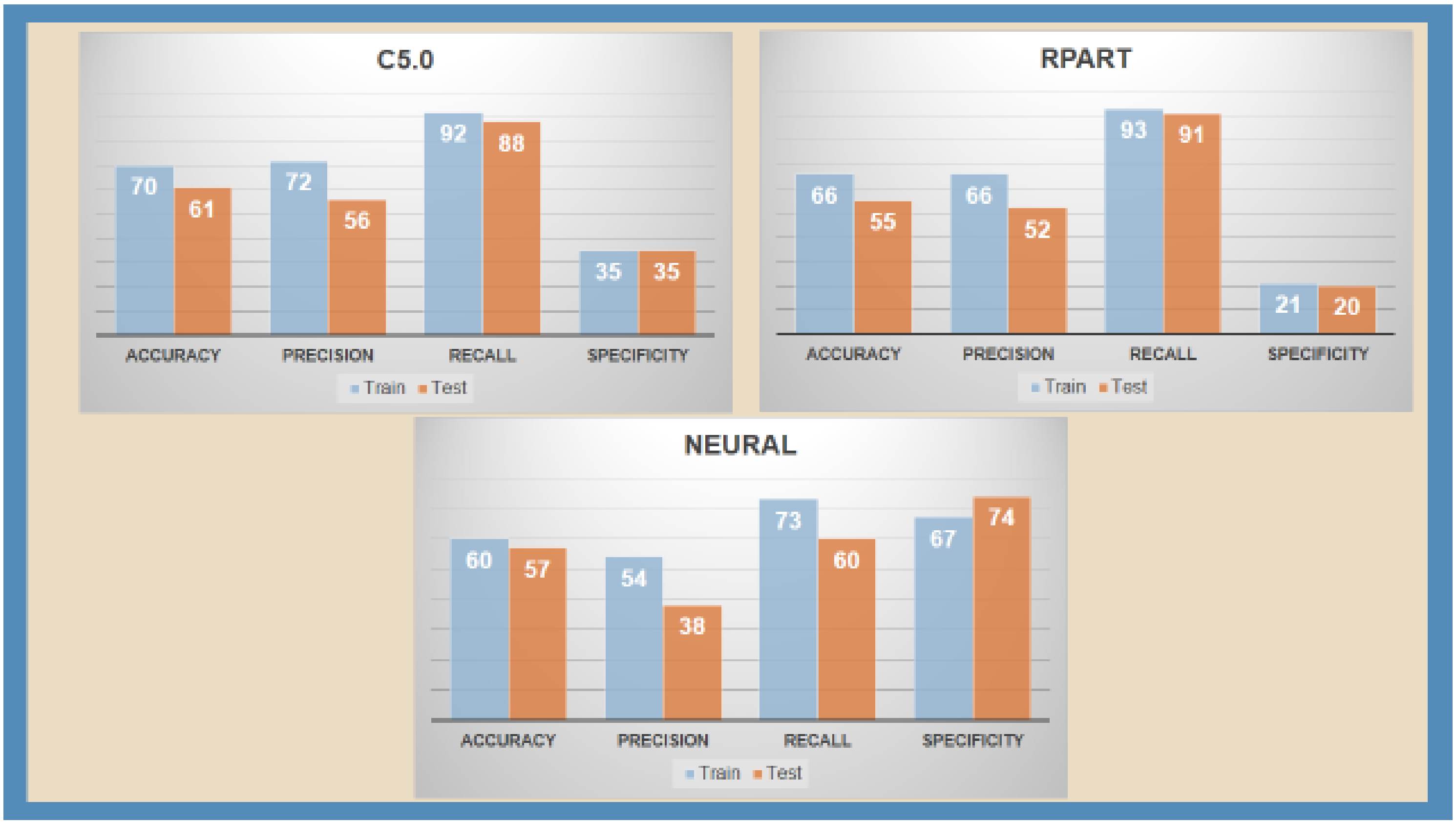
LogReg <- glm(formula = target ~ ActGPI + QTY + Age + PurchasedBy + Pharmacy, family = binomial, data = train)

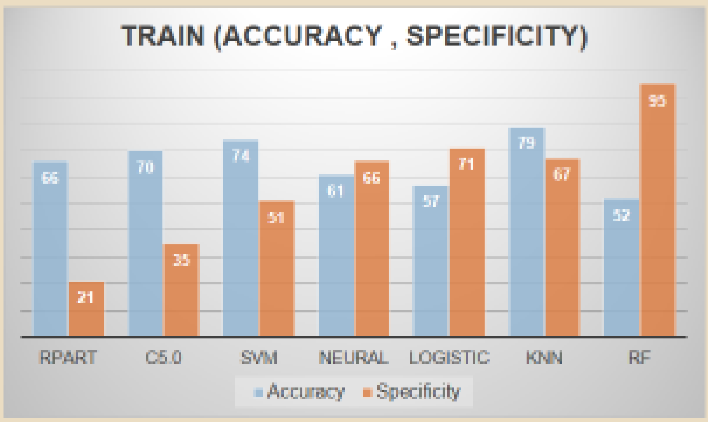

RANDOM FOREST
52
56
26
14
90
76
95
95
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test

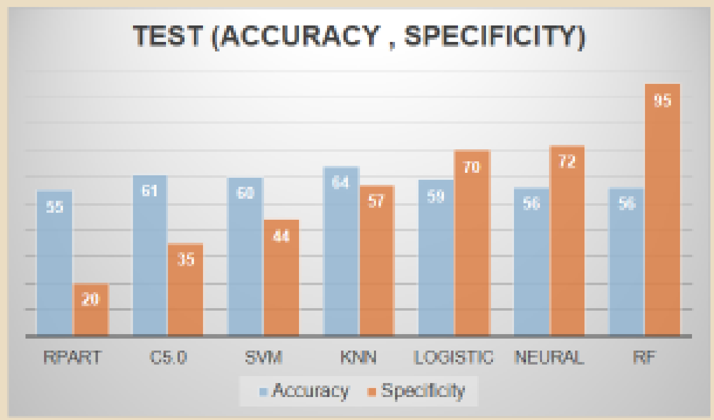

rf
t
Pharmacy
AmountPaid
QTY
Age
RouteOfAdmin
PurchasedBy
Sex
State
MailRetail
0
100
200
300
400
500
600
MeanDecreaseGini

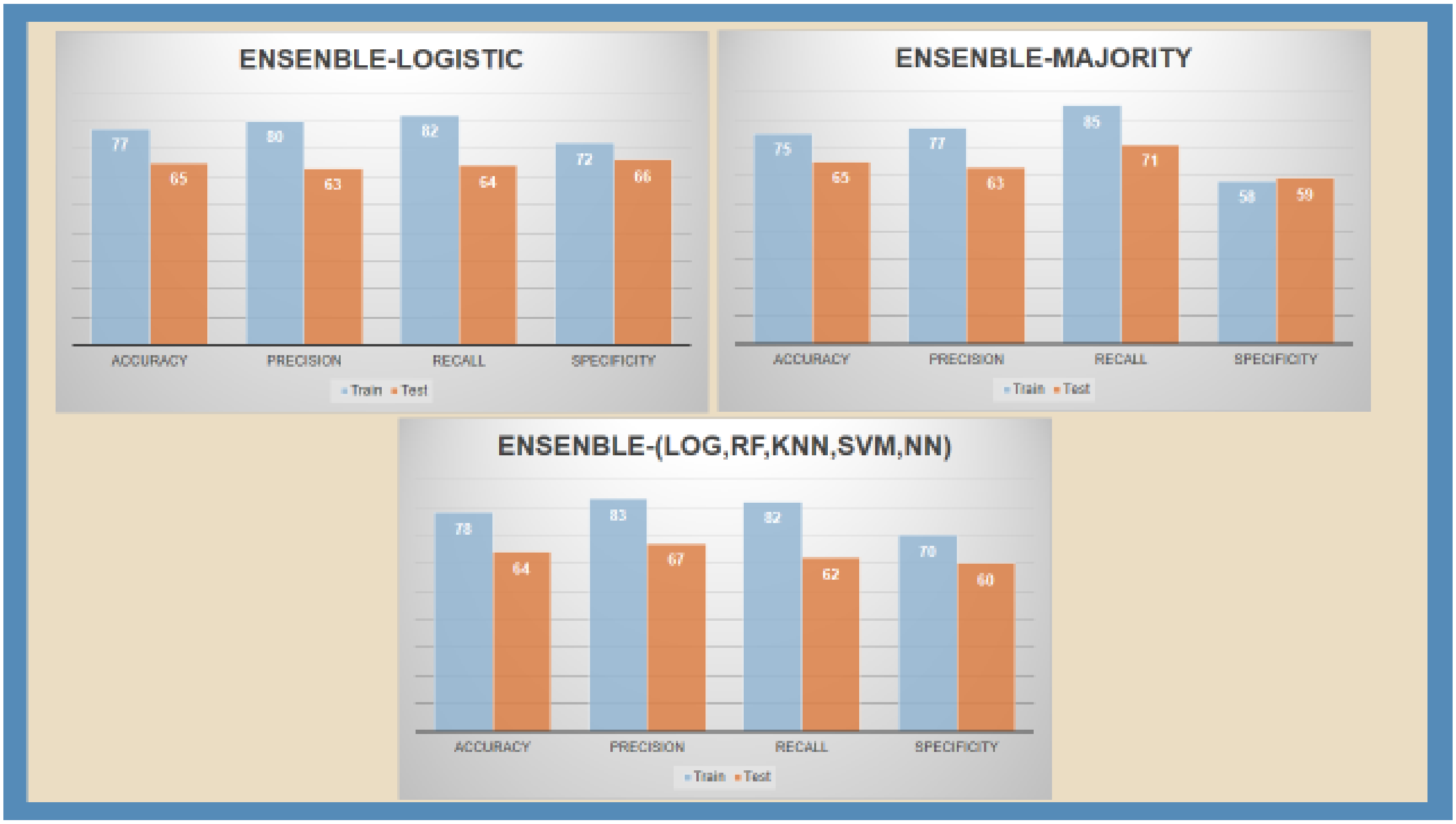

KNN
79
64
81
61
86
70
67
57
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test

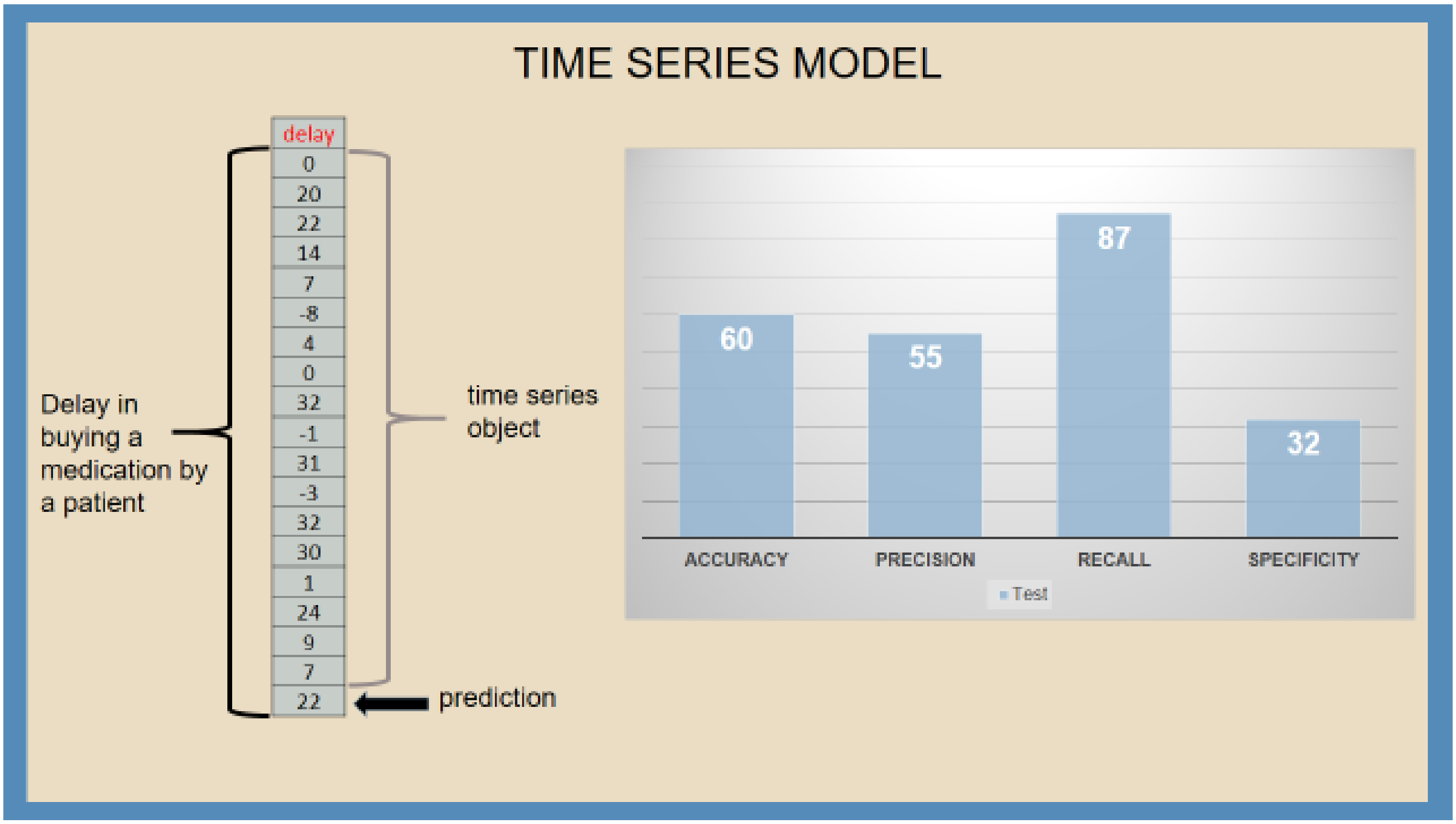

SVM
74
60
74
57
88
77
51
44
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test


C5.0
70
61
72
56
92
88
35
35
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test
RPART
66
55
66
52
93
91
21
20
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test
NEURAL
60
57
54
38
73
60
67
74
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test

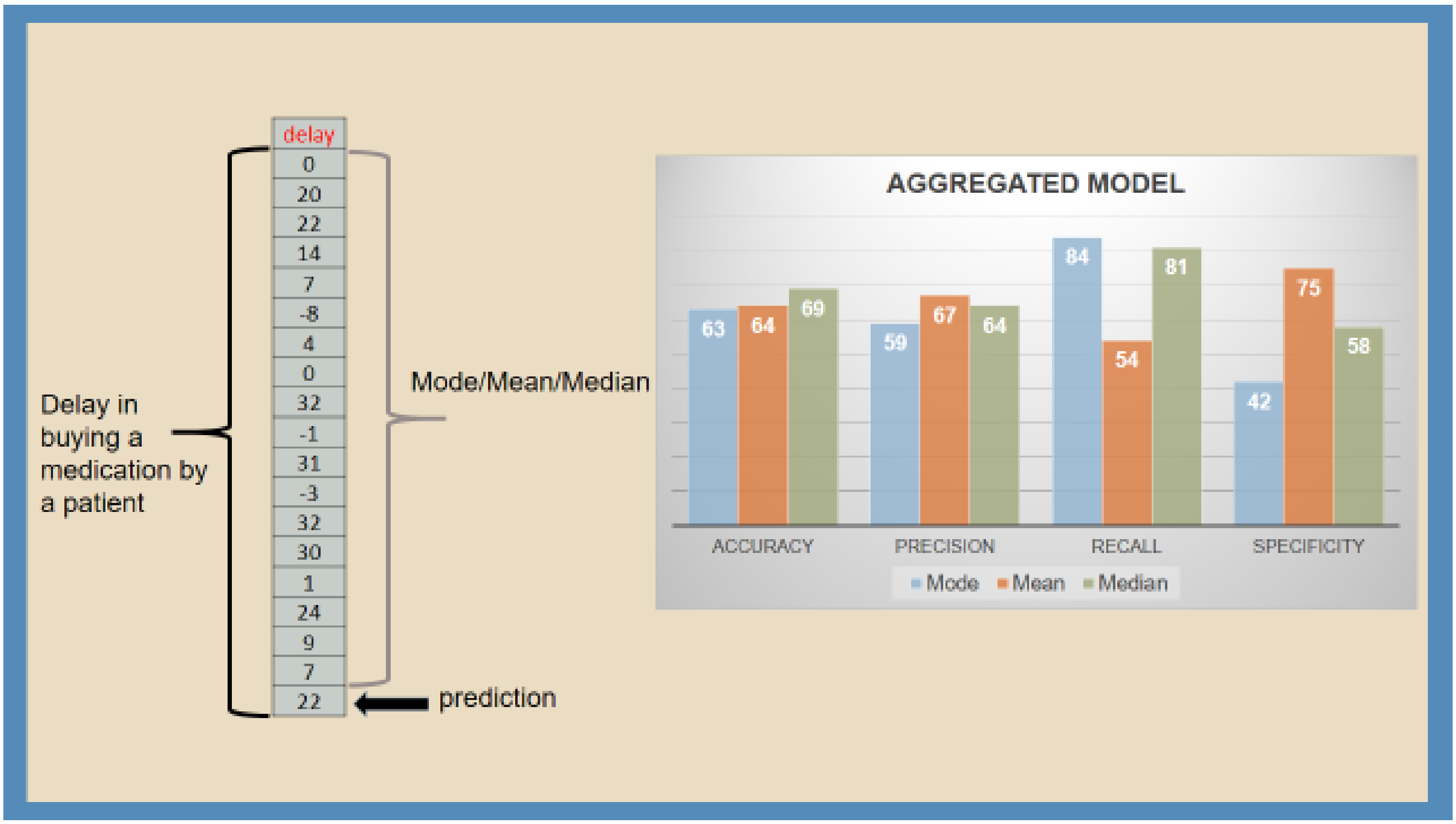

TRAIN (ACCURACY , SPECIFICITY)
66
21
70
35
74
51
61
66
57
71
79
67
52
95
RPART
C5.0
SVM
NEURAL
LOGISTIC
KNN
RF
Accuracy
Specificity


TEST (ACCURACY , SPECIFICITY)
55
20
61
35
60
44
64
57
59
70
56
72
56
95
RPART
C5.0
SVM
KNN
LOGISTIC
NEURAL
RF
Accuracy
Specificity

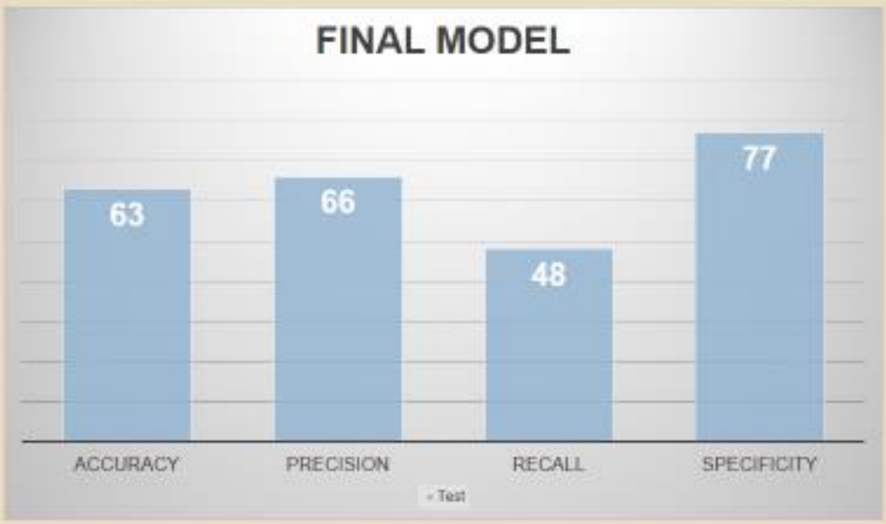

ENSENBLE-LOGISTIC
77
65
80
63
82
64
72
66
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test
ENSENBLE-MAJORITY
75
65
77
63
85
71
58
59
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test
ENSENBLE-(LOG,RF,KNN,SVM,NN)
78
64
83
67
82
62
70
60
ACCURACY
PRECISION
RECALL
SPECIFICITY
Train
Test



TIME SERIES MODEL
delay
0
20
22
14
7
-8
4
0
32
-1
31
-3
32
30
1
24
9
7
22
Delay in buying a medication by a patient
time series object
prediction
60
55
87
32
ACCURACY
PRECISION
RECALL
SPECIFICITY
Test



delay
0
20
22
14
7
-8
4
0
32
-1
31
-3
32
30
1
24
9
7
22
Delay in buying a medication by a patient
Mode/Mean/Median
prediction
AGGREGATED MODEL
63
64
69
59
67
64
84
54
81
42
75
58
ACCURACY
PRECISION
RECALL
SPECIFICITY
Mode
Mean
Median



Logistic + RF + KNN + NN + Mean

MODEL EVALUATION	TRAIN				TEST			
	Accuracy	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity
LOGISTIC	57	73	48	71	59	60	50	70
RANDOM FOREST	52	26	90	95	56	14	76	95
KNN	79	81	86	67	64	61	70	57
SVM	74	74	88	51	60	57	77	44
C5.0	70	72	92	35	61	56	88	35
RPART	66	66	93	21	55	52	91	20
NEURAL NETWORKS	60	54	73	67	57	38	60	74
Ensemble-logistic (all models)	77	80	82	72	65	63	64	66
Ensemble-majority ranking (all models)	75	77	85	58	65	63	71	59
TIME-SERIES	-	-	-	-	60	55	87	32
AGGREGATED-MODE	-	-	-	-	63	59	84	42
AGGREGATED-MEAN	-	-	-	-	64	67	54	75
AGGREGATED-MEDIAN	-	-	-	-	69	64	81	58
FINAL					63	67	50	77