



PREDICTING SUCCESS WITH SIMPLYX

PETER KINDER & MICHAEL GHATTAS

ALTERYX: DATA PREP & ANALYTICS

The screenshot displays the Alteryx Designer interface with a project titled "Customer Campaign Analytics.yxz". The interface is divided into several sections:

- Toolbar:** Includes File, Edit, View, Options, Help, and various tool icons for In/Out, Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Spatial, Interface, Data Investigation, Predictive, AB Testing, Time Series, Predictive Grouping, Prescriptive, Connectors, Address, and more.
- Browse (157) - Configuration:** Shows a scatter plot of "Score_Yes" (ranging from 0 to 1) against an unnamed column (ranging from 0 to 1000). Below the plot is a histogram of "Score_Yes" with a single bar at 0. A table provides summary statistics for "Score_Yes":

Data Type	Double
Size	8
Non-Nulls	1245
Nulls	0
Minimum	0.142927381036451
Maximum	0.99999999920593
Average	0.237751003996273
Standard Deviation	0.162859151408325
Variance	2.65231031974396e-002
- Workflow Area:** A complex data pipeline is shown, starting with "Input Data" (Transactions, Customer Database, Stores) and moving through "Data Preparation" (Product_Categor, FORMULA & CALCULATE FIELDS, SUMMARIZE, FIELD SELECTION, TEXT PARSING), "Data Blending" (JOIN MULTIPLE, JOIN, CREATE POINTS FROM CUSTOMER ADDRESS, CREATE POINTS FROM LAT/LON), "Geographic Analysis" (Trade Area), "Predictive Analytics" (LOGISTIC, Model Building, Output, Interactive), and finally "Share Analyses / Data" (BROWSE, Local Output). A "Model Deployment" section is also present.
- Results - Browse (157) - Input:** A table showing the first 7 records of the input data:

Record #	Customer_ID	Spend	Transactions	Customer.ID	Store.Number	Customer.Segment	Responder	Name	Address	Postcode	Lat	Lon	ID	Cty	R
1	2391	537.65	1	2391	100	Home Office	No	TONY HEFLIN	9356 W 107TH MEWS	C87 4PG	52400681	026571	100	Ely	C87
2	2404	1156.91	1	2404	100	Corporate	No	VANEK FLINT	12650 W 64TH AVE	C87 4AL	52400024	0259752	100	Ely	C87
3	2423	22.62	1	2423	100	Corporate	Yes	PHILIP MUMFORD	2368 E FLOYD PL	C87 4LY	523399372	0265525	100	Ely	C87
4	2431	1087.36	1	2431	100	Small Business	Yes	ASHENAFL BEENTI	787 DILLON WAY	C87 4LY	523399372	0265526	100	Ely	C87
5	2442	485.8	1	2442	100	Home Office	No	CATHERINE DALY	6350 DEFRAME WAY	C86 1BF	52405586	0266035	100	Ely	C87
6	2447	159.31	1	2447	100	Corporate	Yes	PEGGY EYEMANN	207 TROY ST	C87 4NY	52399212	0265884	100	Ely	C87
7	2450	0	1	2450	100	Corporate	No	ERIC OLSON	5605 S BUCHANAN ST	C87 4HE	0	0	100	Ely	C87

A GLOBAL CUSTOMER BASE

~7,000 CUSTOMERS ACROSS DIVERSE INDUSTRIES

TOP LINE GROWTH		EFFICIENCY GAINS		FAST WORKFORCE RESKILLING	
RETAIL	FOOD SERVICES	CONSUMER PRODUCTS	TELCOM/CABLE	MEDIA	PROFESSIONAL SERVICES
MUJI 無印良品   GEMS   Al-Futtaim   GROUPON   IKEA  	kikkoman®  Campbell's   coles  Sainsbury's  Nestle  Kellogg's  DANONE  KraftHeinz  THE ARART HEINZ COMPANY	BAYER  ESTEE LAUDER  SHISEIDO  P&G  Unilever  Johnson & Johnson 	COX  iHeart MEDIA  TELUS  Claro  vodafone  swisscom  Bell  telenor  	GIARTE  NETFLIX  ITV  FOX  Tribune direct  Omnicom  UNIVERSAL UNIVERSAL MUSIC GROUP  	KKR  HAKUHODO  HARTE HANKS  High Point  CLARIANT  Marcellin Michelin  TATA  Tata Consultancy Services  BDO 
AMBIT ENERGY  PETROBRAS DISTRIBUIDORA S.A.   edf  WorldFirst   DELTA  Chevron   Sempra Energy  TEPCO Energy Partner  	Abbott  SHARP  日本生命    Unum  novonordisk   Novo Nordisk  National Skin Kerting Cancer Center  	SIEMENS  SAINT-GOBAIN  Hallmark  Audi   Ingersoll Rand.  SABEMI  AIRBUS  BRIDGESTONE 	First Commonwealth  HSBC  [B] ³  VISA  WELLS FARGO  	ANA  MGM RESORTS  QATAR  Marriott   TOLL  TotalCaribbean  expedia group 	IBM  Microsoft  facebook  amazon services   FUJITSU  intuit  SAMSUNG ELECTRONICS   expedia group  
alteryx The Thrill of Solving					Note: Customers as of Q3 2020

38% OF THE GLOBAL 2000

ALTERYX CERTIFICATIONS

Alteryx Designer Core



Free

Beginner

2 Hours

80 Questions

Prep Guide

REGISTER



Alteryx Designer Advanced



Free

Advanced

2 Hours

44 Questions

Prep Guide

REGISTER



Alteryx Designer Expert



USD 150

Expert

2 Hours

4-5 Questions

REGISTER



Alteryx Partner Certification



Free

Partner

2 Hours

80 Questions

REGISTER



SIMPLYX

- DESIGNED TO FACILITATE ALTERYX DESIGNER CORE CERTIFICATION
- COURSEWARE BUILT ATOP ALTERYX
- DEMONSTRATED TO SIGNIFICANTLY BOOST PASS RATE



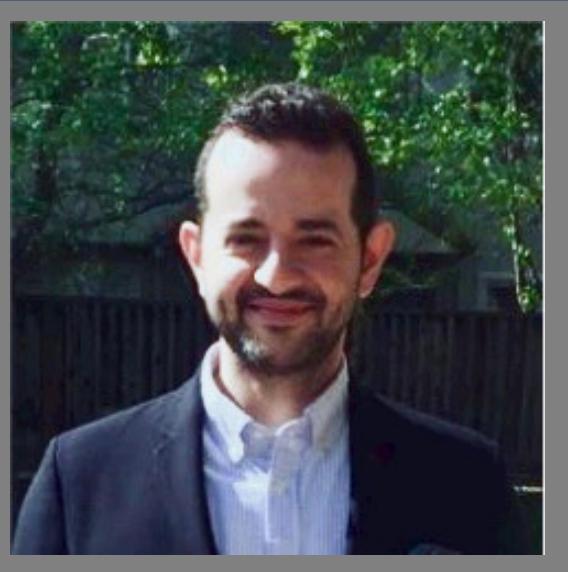
SIMPLYX EXAMPLE

The screenshot shows the Alteryx Designer interface with the title "Purple Belt.yxmd". The top menu bar includes File, Edit, View, Options, Help, and a search bar. The toolbar contains various tools like Favorites, In/Out, Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Spatial, Machine Learning, Text Mining, Computer Vision, Interface, Data Investigation, Predictive, AB Testing, Time Series, and Predictive Grid. The left sidebar shows "Workflow - Configuration" with tabs for Canvas, Workflow, Runtime, Events, and Meta Info. The main canvas area is titled "Exercise 15" and contains a workflow diagram. The diagram starts with two "Input Data" tools (represented by blue book icons) on the left. A red square box highlights a "Summarize" tool (red icon with a sigma symbol) located in the middle section of the workflow. The bottom section of the canvas shows a "Results - Workflow - Messages" panel with a summary of errors, warnings, and info messages. The status bar at the bottom right indicates "Last Run" and "Configuration".

EXERCISE 15

Your bosses at Ralphie Inc. want a summary of the total sales for each product. They want a table with the Product Name as well as Total Sales. Additionally, they are not interested in any of the products that have been discontinued. They also wanted to remind you not to forget to consider and discounts (discount information is provided in decimal format. i.e. 0.15 = 15%)

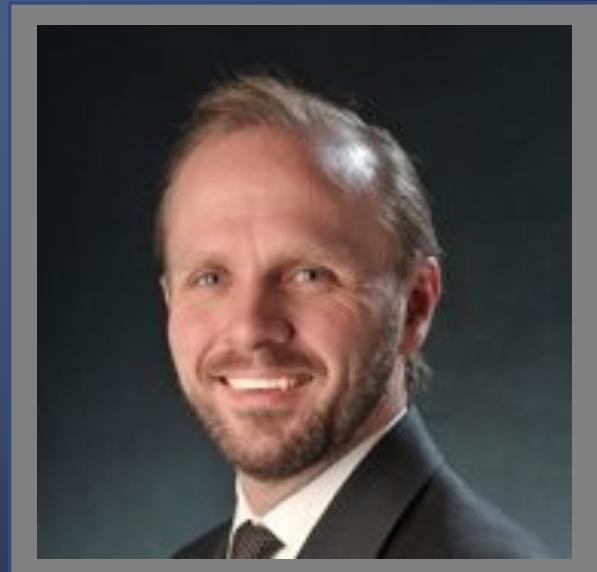
THE TEAM



MICHAEL GHATTIS



PETER KINDER

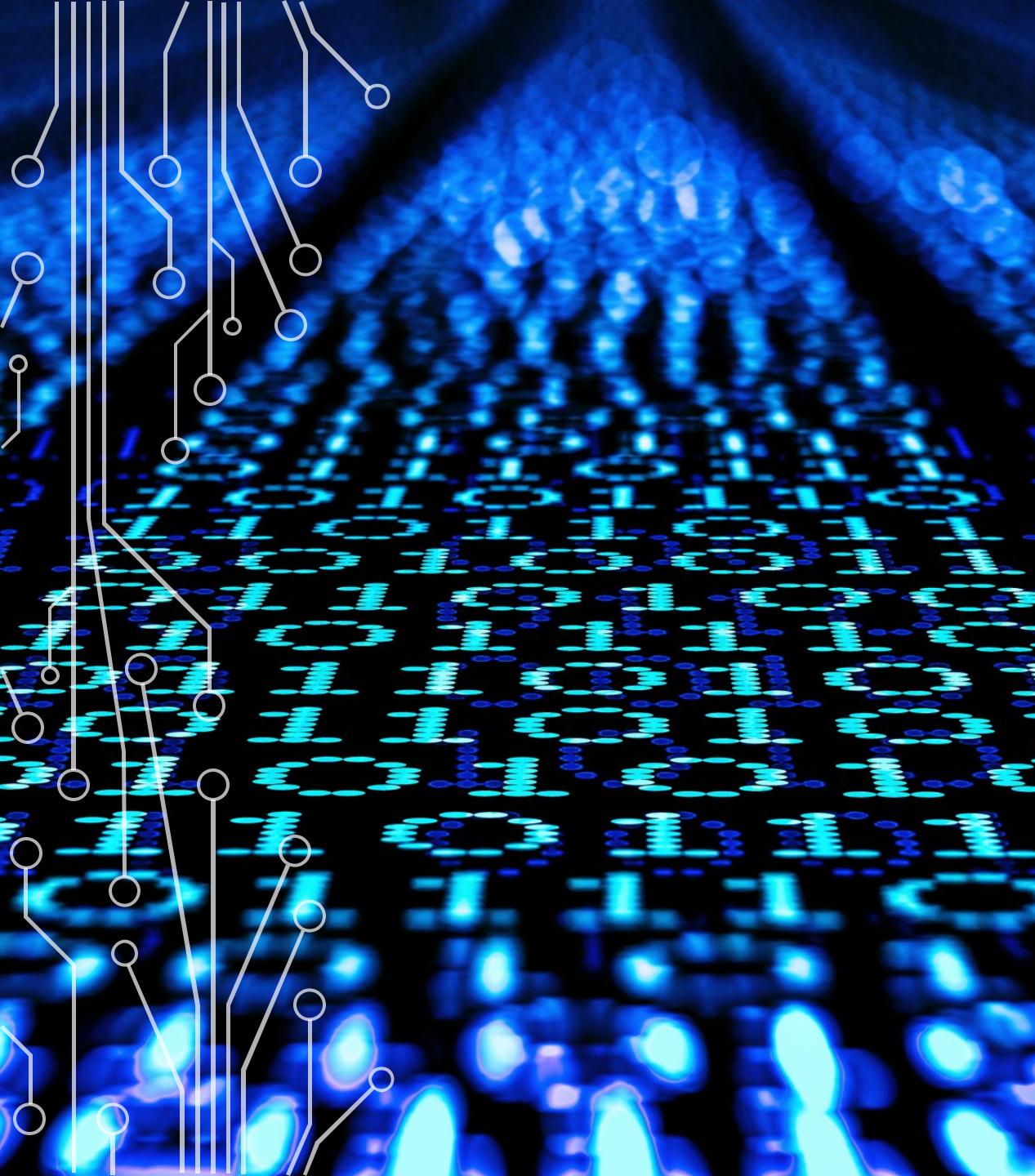


KAI LARSEN



THE PROBLEM

DEVELOP A MODEL TO
PREDICT USER SCORES
ON THE ALTERYX
DESIGNER CORE
CERTIFICATION WITH
DATA OBTAINED FROM
SIMPLYX USE



THE DATA



- 100K+ USER LOG FILES
 - USER ID
 - PROBLEM #
 - BELT (SECTION) #
 - CORRECT SCHEMA (T/F)
 - CORRECT ANSWER (T/F)
 - TIMESTAMP
- 206 SIMPLYX USERS' DESIGNER CORE CERTIFICATION SCORES

DATA PREP

- 6 BELT (SECTION) % SCORES
- 148 ONE HOT ENCODED VALUES
 - EACH PROBLEM SOLVED OR NOT
- 5 USAGE SUMMARY VALUES
 - TOTAL ATTEMPTS
 - TOTAL SESSIONS (1 HOUR SEPERATION)
 - AVG CORRECT PER SESSION
 - AVG ATTEMPTS PER SESSION
 - AVG SECONDS PER SESSION

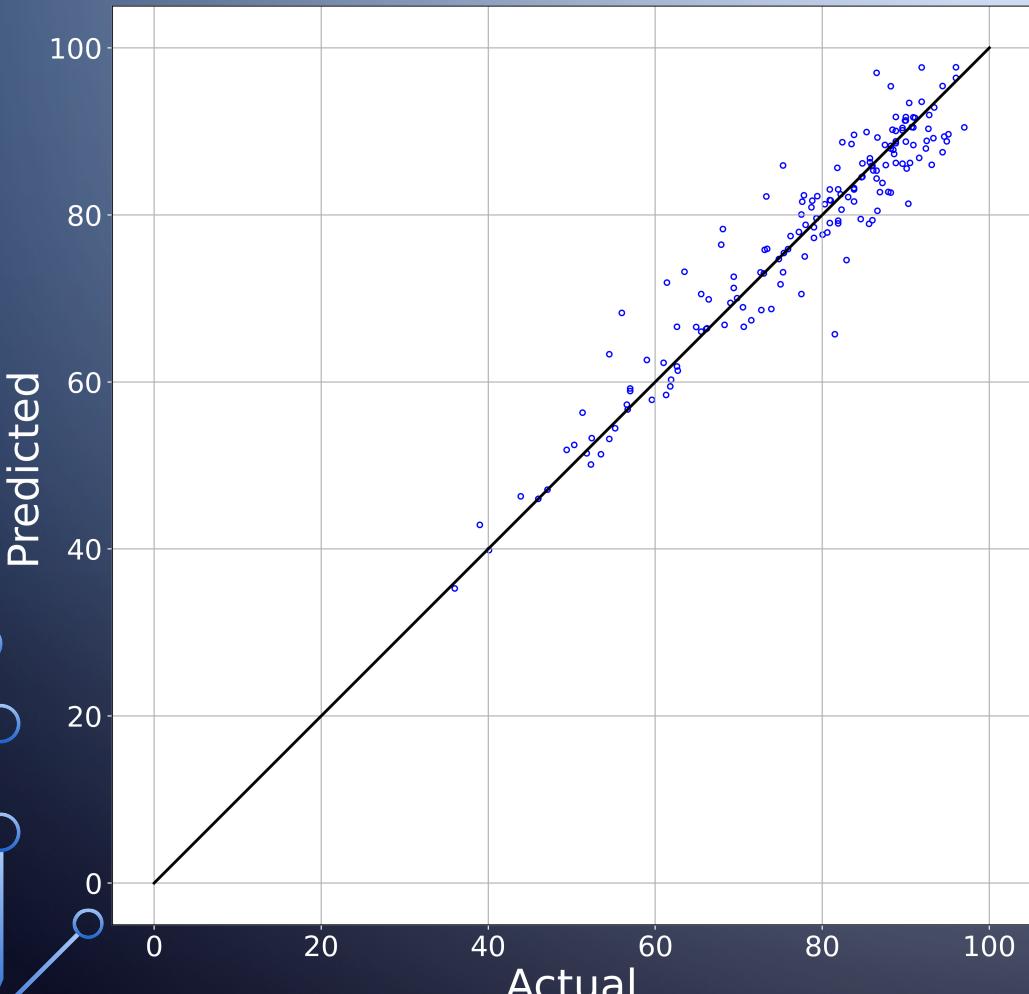


STEPWISE MLR: TRAINING SET

AIC

R2: 0.943

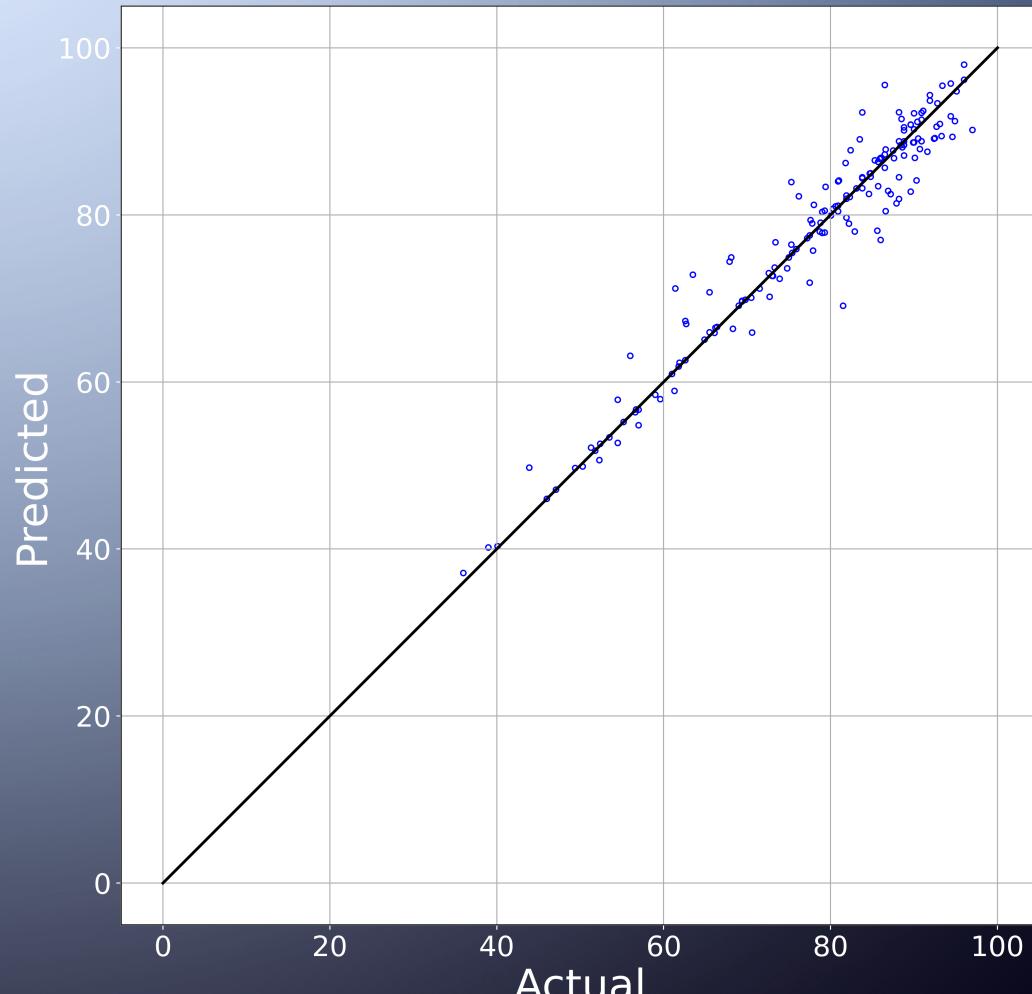
RMSE: 3.294



BIC

R2: 0.9152

RMSE: 4.019

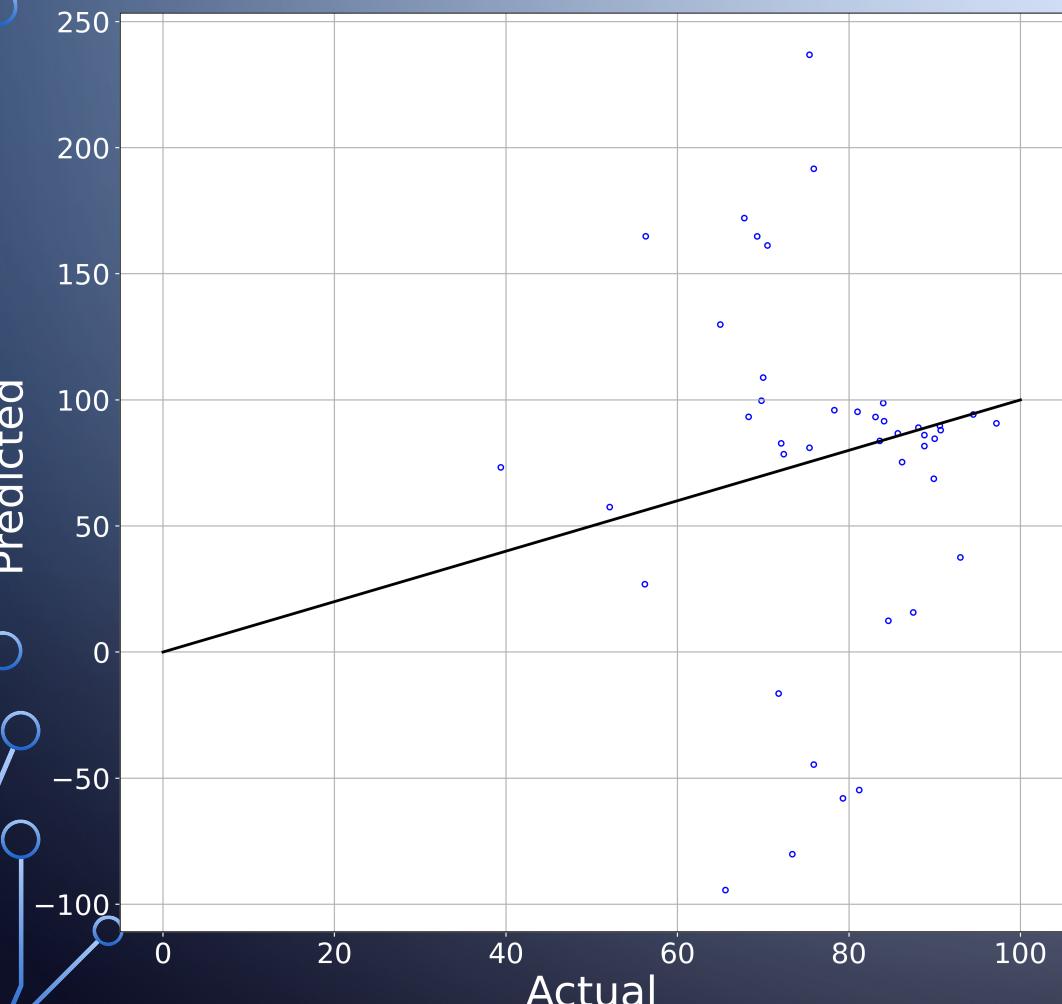


STEPWISE MLR: TEST SET

AIC

MAE: 48.685

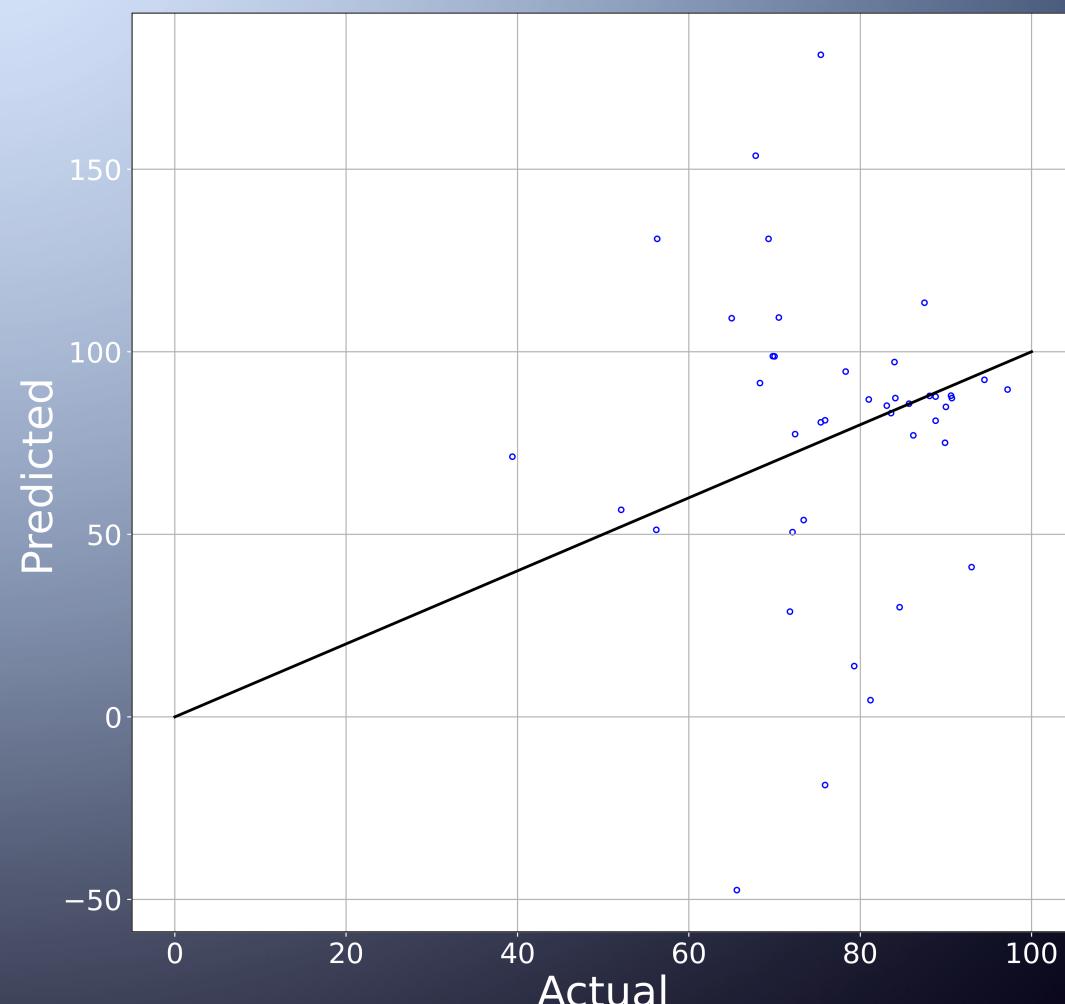
RMSE: 71.183



BIC

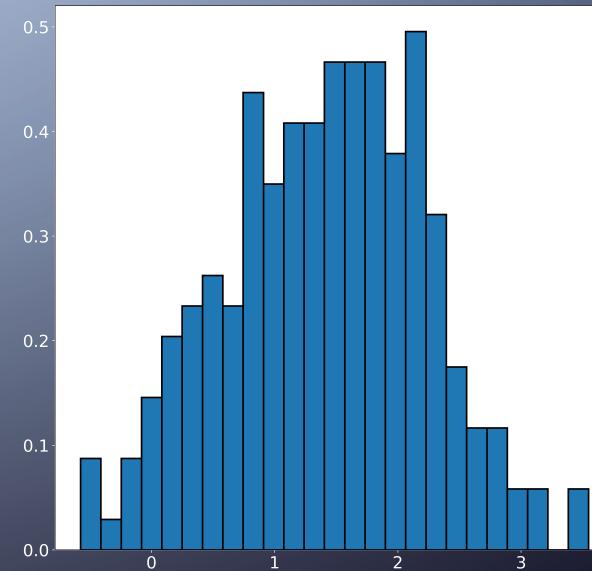
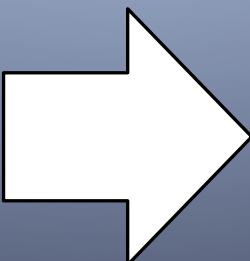
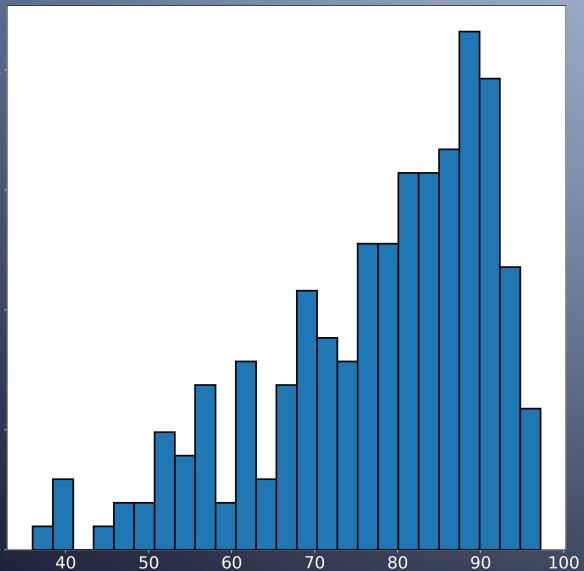
MAE: 28.806

RMSE: 42.709



TRANSFORMING THE RESPONSE

- DIVIDED BY 100
 - CONVERTED FROM PERCENT TO RATE
- PERFORMED LOGIT TRANSFORMATION
 - $\ln(Y / (1 - Y))$
 - BOUNDED THE PREDICTION
 - NORMALIZED DISTRIBUTION

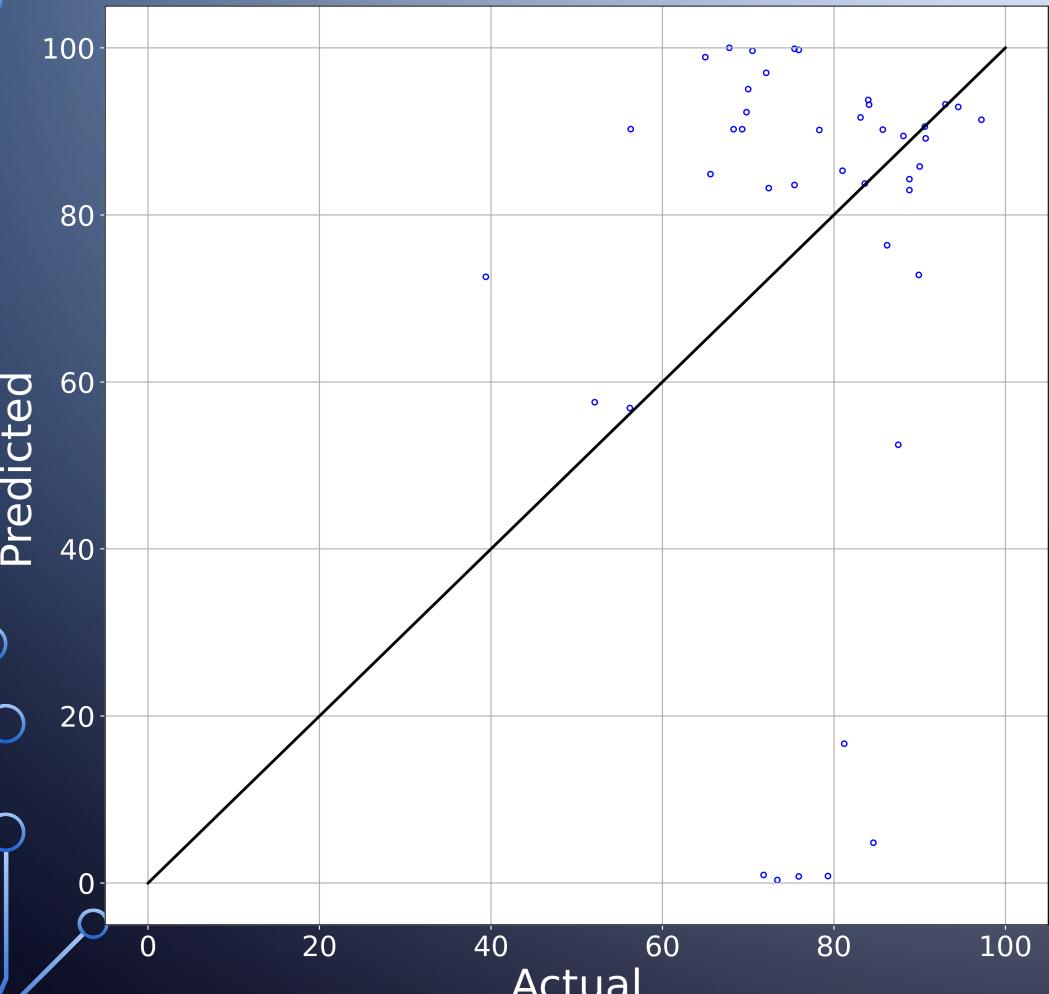


STEPWISE MLR: TEST SET

AIC

MAE: 22.558

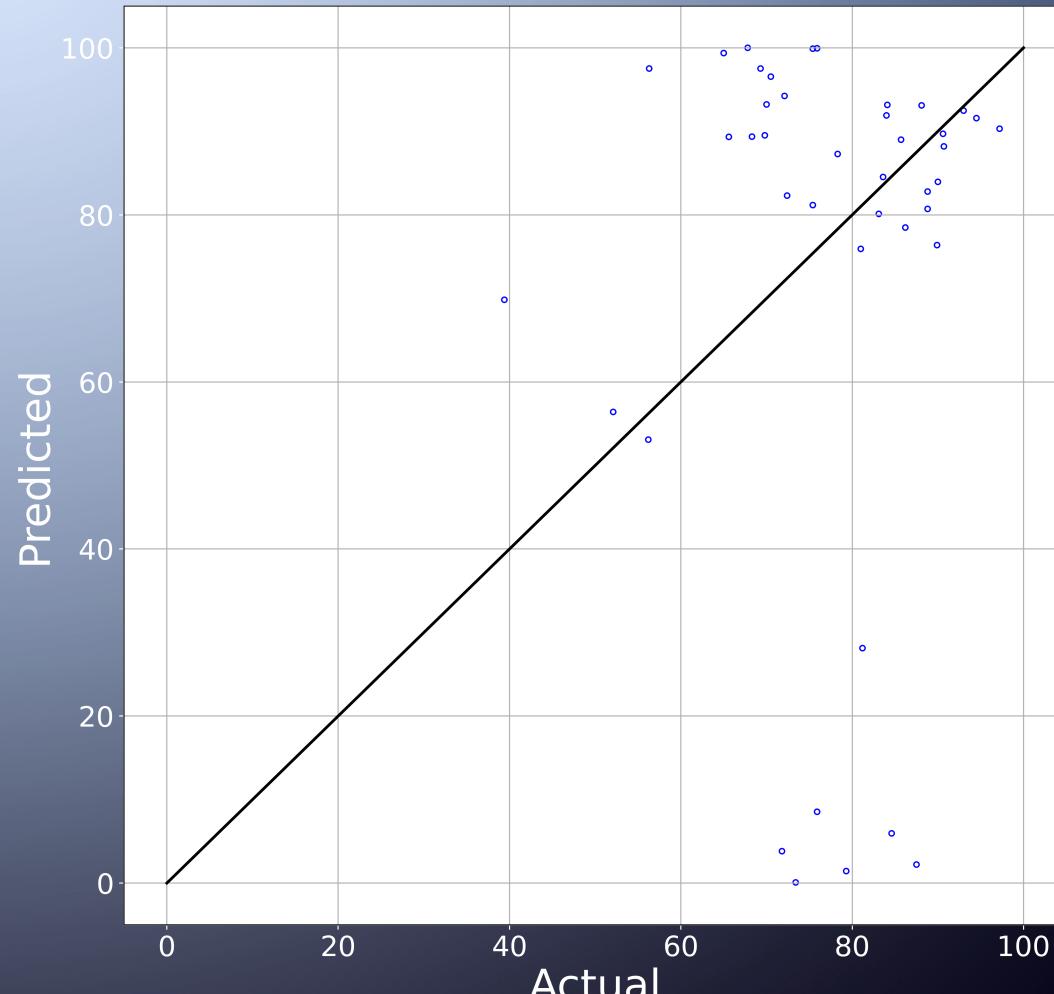
RMSE: 32.567



BIC

MAE: 23.235

RMSE: 33.697



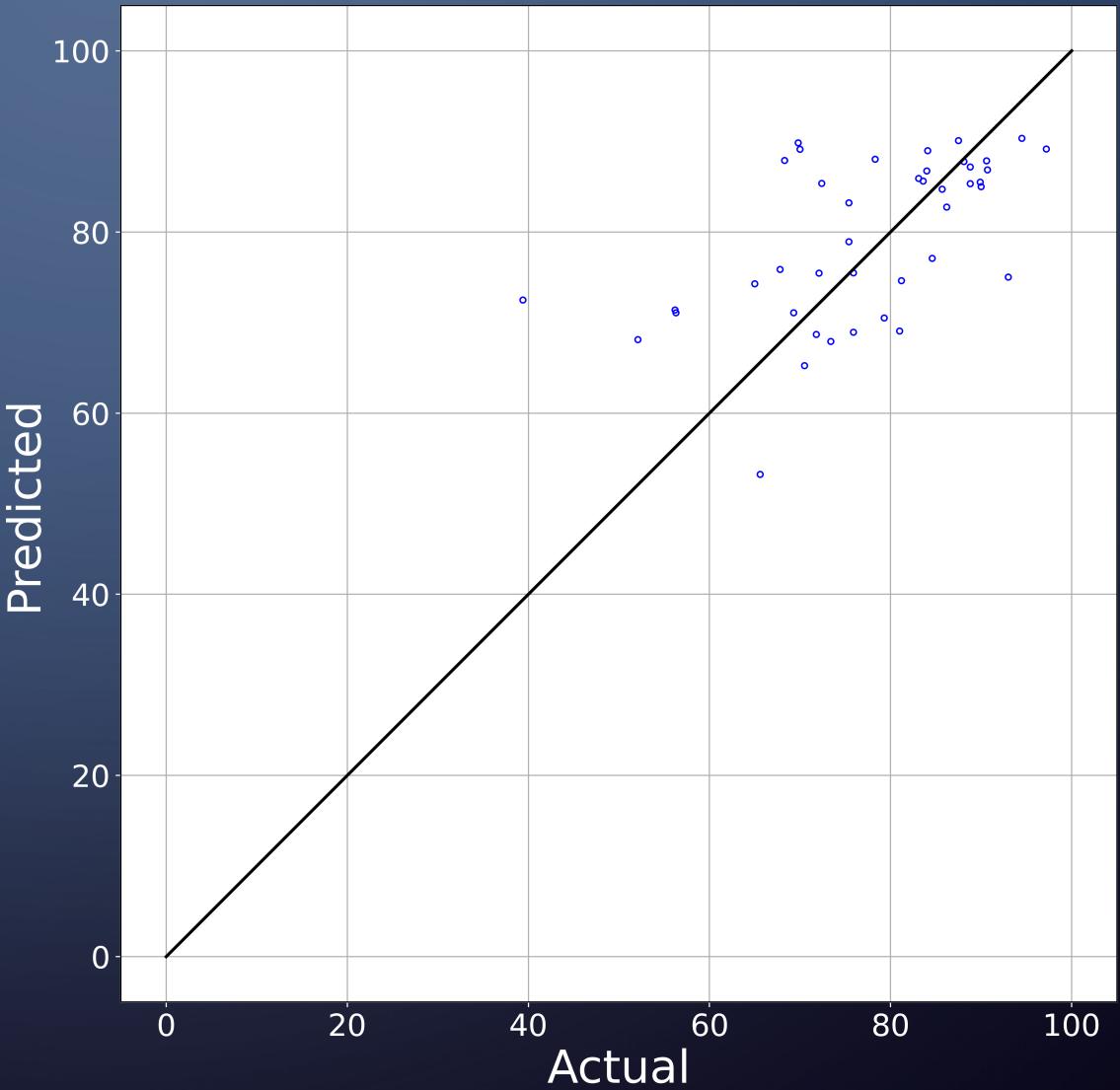


ADDRESSING OVERTFITTING

- REGULARIZATION
 - DIMENSION REDUCTION
 - CROSS VALIDATION
 - ENSEMBLES
- 

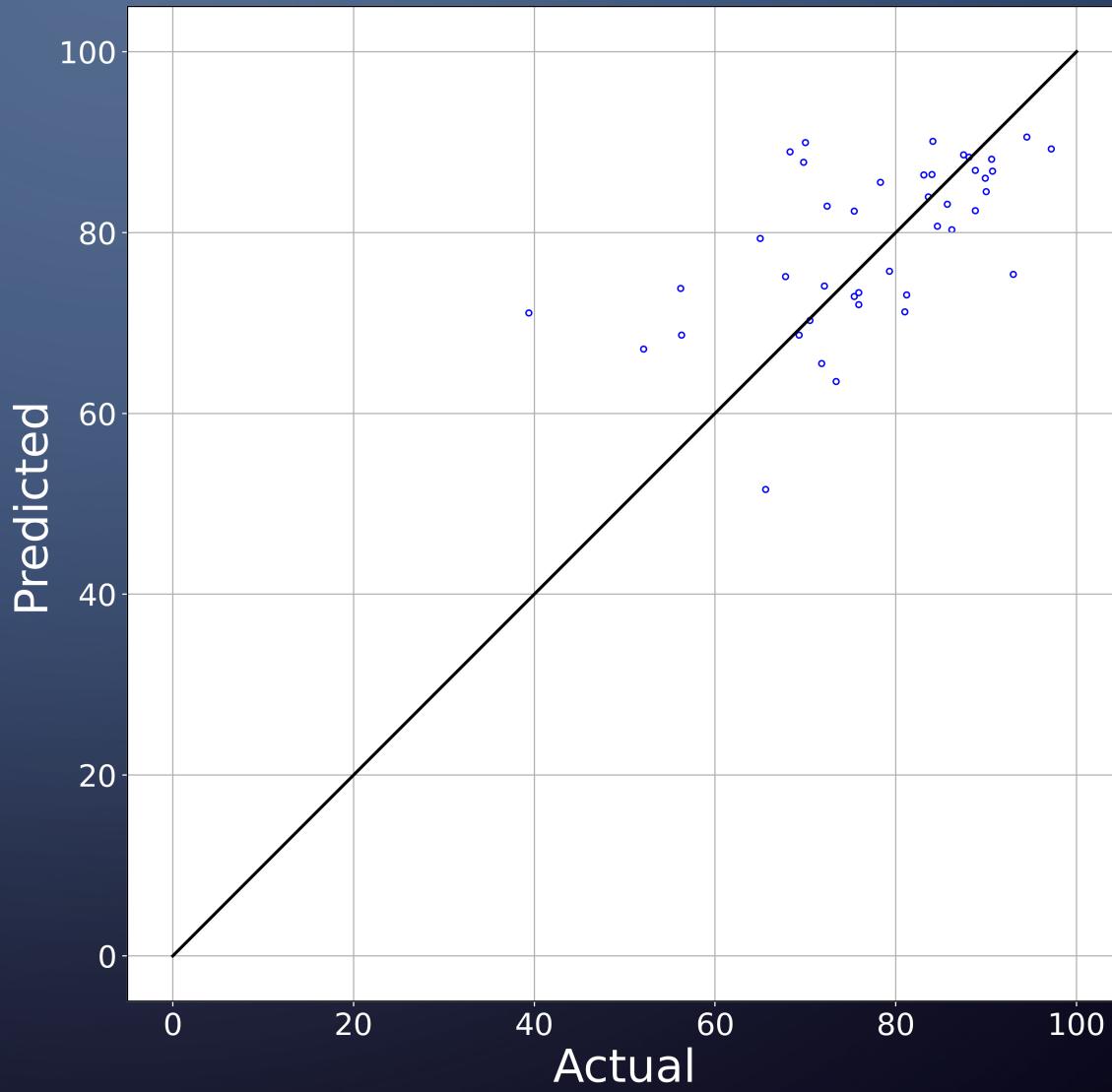
RIDGE REGRESSION

- L2 REGULARIZATION
- ADDRESSES MULTICOLLINEARITY
- GRID SEARCH FOR ALPHA
 - MINIMIZE RSS VS SUM OF SQUARES COEFS
 - COEFS RARELY ZERO
- MAE: 7.948
- RMSE: 10.455



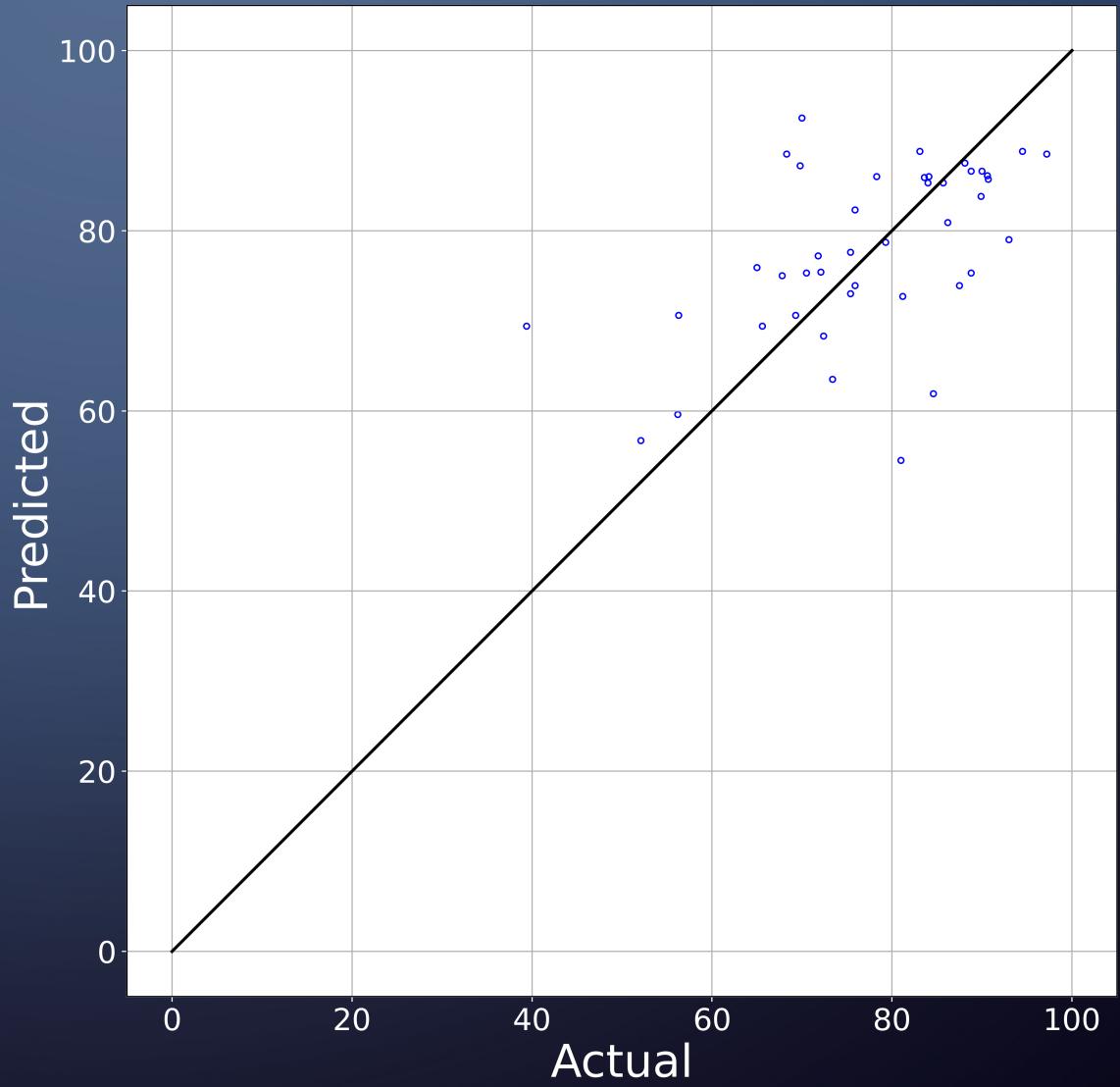
LASSO

- L1 REGULARIZATION
- ADDRESSES MULTICOLLINEARITY
- GRID SEARCH FOR ALPHA
 - MINIMIZE RSS VS MAGNITUDE OF COEFS
 - COEFS OFTEN ZERO
- MAE: 7.722
- RMSE: 10.306



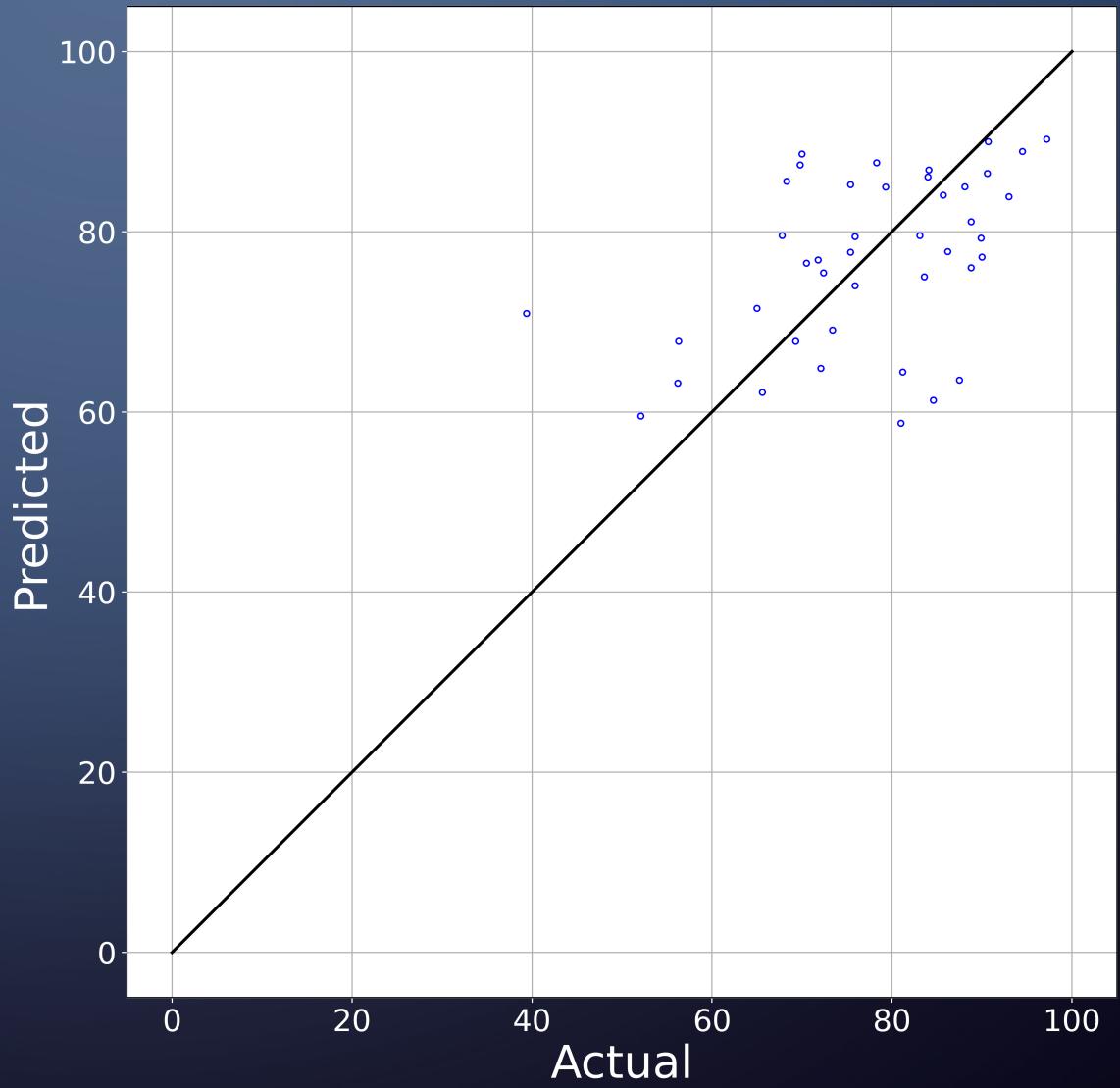
BOOSTING

- LEARNS FROM PREVIOUS TREES
- GRID SEARCH FOR
 - # OF TREES
 - LEARNING RATE
 - DEPTH
 - % OF FEATURES
- MAE: 8.007
- RMSE: 10.874



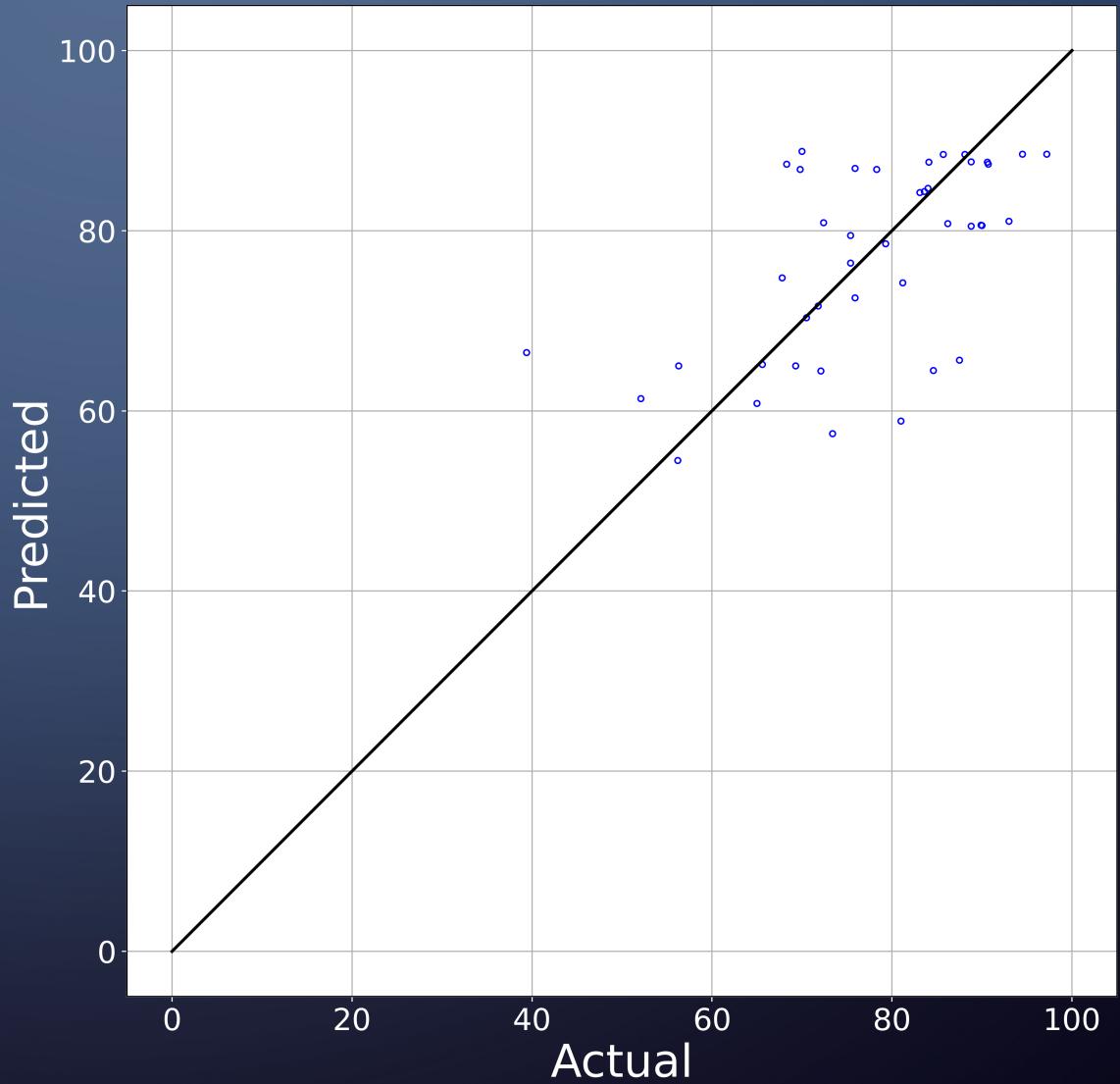
BAGGING

- BOOTSTRAP WITH REPLACEMENT
- GRID SEARCH FOR
 - # OF TREES
 - % OF SAMPLES
 - % OF FEATURES
- MAE: 9.032
- RMSE: 11.462



RANDOM FOREST

- SPLIT ON SUBSET OF FEATURES IN EACH SPLIT
- GRID SEARCH FOR
 - # OF TREES
 - DEPTH
 - % OF FEATURES
- MAE: 7.750
- RMSE: 10.454

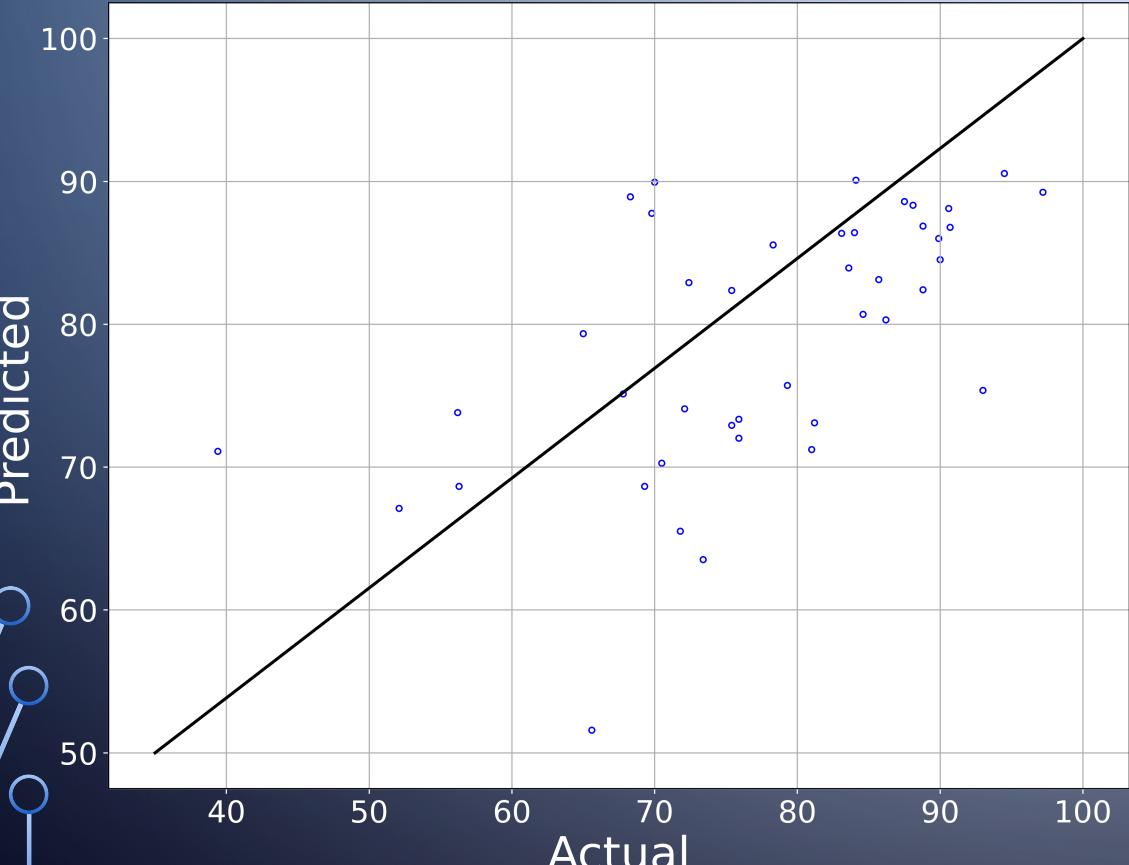


TIE BREAKER

LASSO

MAE: 7.722

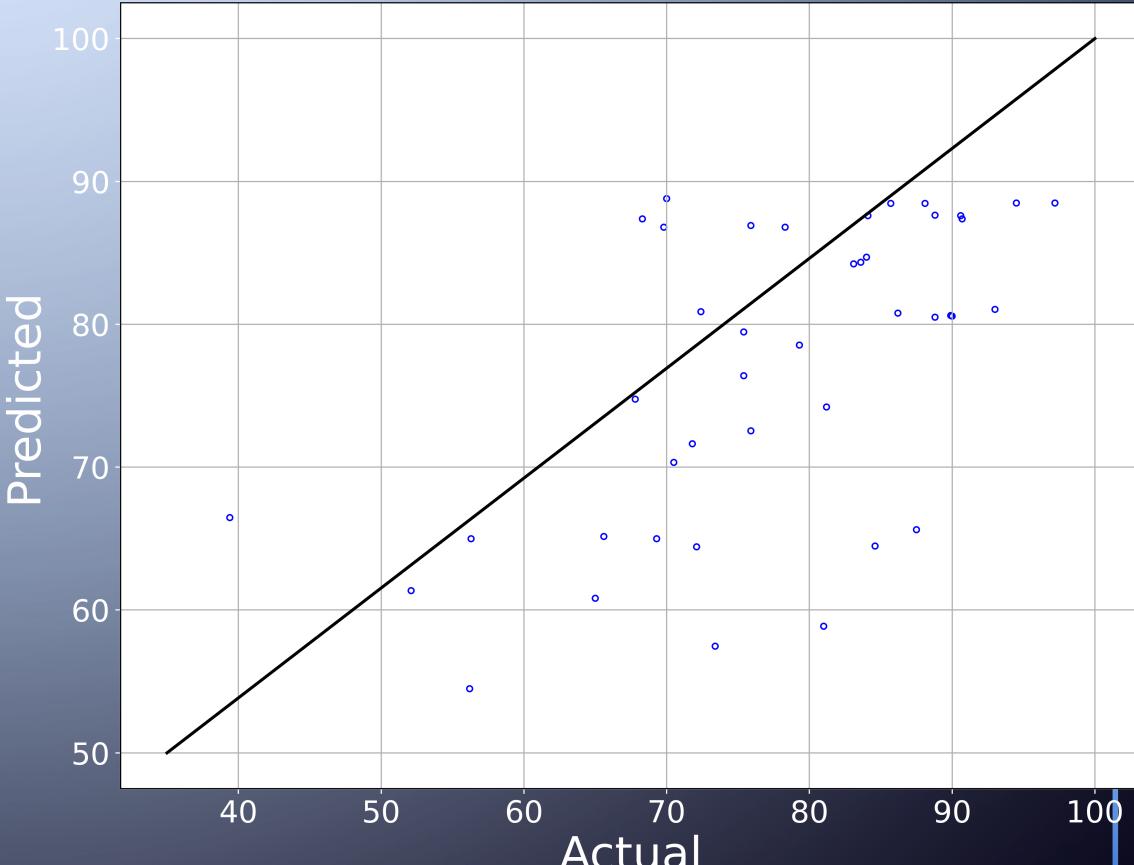
RMSE: 10.305



RANDOM FOREST

MAE: 7.750

RMSE: 10.454



CONCLUSION

- KAI (DE) WAS HAPPY WITH THE RESULTS
- WILL BE DEPLOYED FOR FALL 2023 SEMESTER
- MORE DATA WOULD BE HELPFUL
 - # OF SAMPLES
 - STUDENT EDUCATIONAL HISTORY (PERHAPS A SURVEY?)
- HYPOTHESIZE THAT SOME STUDENTS MAY COPY OTHERS
 - MAKES PREDICTION DIFFICULT



QUESTIONS?