Predicting High School Track Times

Dana Lindquist
Project Luther
October 12, 2018

High school track statistics are collected and stored at www.athletic.net

Given:

    9th grade PR (Personal Record)

    10th grade PR

    11th grade PR

Predict:

    12th grade PR

Athletic.net *Track & Field*

Q Search   ❓ Get Help   ☆ Favorites   👤 Dana

Records ▾   Rankings ▾   Custom Lists ▾   ⭐ Reports ▾   Training Log   Photos

View Athletic.net Ad Free

👤 **Patrick Gibson** Track & Field Profile   [Claim this Athlete]   [Go To *XC* Profile]

All   High School   ✎ Make a Correction

2013 Outdoor Season Squalicum HS   12th Grade

800 Meters

2	1:59.1h	Mar 15	Squalicum, Bellingham, and Sehome	V F
1	2:06.60	Mar 28	Lynden vs Squalicum	V F
3	1:58.13 PR	Apr 19	NWC Championship Meet	V F

1600 Meters

1	4:36.24	Mar 28	Lynden vs Squalicum	V F
1	4:16.2h	Apr 6	Birger Solberg Invitational	V F
1	4:30.66	May 2	Mt. Baker & Squalicum @ Sehome	V F
1	4:18.66	May 8	NWC 2A Sub District Meet	V F
1	4:18.66	May 10	NWC 2A Sub Distirct Meet	V F
3	4:19.75	May 15	2A District 1 & 2 Bi-District Championship Meet	V F
3	4:19.75	May 17	2A District 1 & 2 Bi-District Championship Meet	V F
1	4:11.85 SR	May 25	2A, 3A, 4A State Championship Meet All Days	2A F

3200 Meters

1	9:15.11	Mar 22	Bedlington Twilight	V F
1	9:58.35	Mar 28	Lynden vs Squalicum	V F
1	9:39.71	Apr 19	NWC Championship Meet	V F
1	9:53.62	May 2	Mt. Baker & Squalicum @ Sehome	V F
1	9:10.40	May 10	NWC 2A Sub Distirct Meet	V F
1	9:05.78	May 17	2A District 1 & 2 Bi-District Championship Meet	V F
2	9:01.80 PR	May 25	2A, 3A, 4A State Championship Meet All Days	2A F

4x400 Relay

8	4:06.8h	Mar 15	Squalicum, Bellingham, and Sehome	V F
			Patrick Gibson, Nick Gallaher, North Bennett, Dillon Rinauro	
1	3:36.0h	May 2	Mt. Baker & Squalicum @ Sehome	V F
			Patrick Gibson, Chais Sugden, Christopher Fredlund, Nick Gallaher	

2012 Outdoor Season Squalicum HS   11th Grade

400 Meters - Relay Split

	56.6h	Mar 16	"City Meet" BHS, SQHS, at Sehome	V F
	54.7h PR	Mar 29	Squalicum w/ Ferndale, Sehome	V F

800 Meters

1	1:59.69	Mar 16	"City Meet" BHS, SQHS, at Sehome	V F
2	1:59.3h SR	Apr 28	Tomahawk T&F Classic	V F

1600 Meters

1	4:39.66	Mar 21	Squalicum w/ Burlington-Edison	V F
1	4:25.0h	Mar 23	Bedlington Twilight	V F
1	4:37.7h	Mar 29	Squalicum w/ Ferndale, Sehome	V F
1	4:18.4h	Apr 20	NWC Championships	V F
1	4:30.0h	Apr 26	Mt. Baker, Squalicum at Blaine	V F
5	4:50.11	May 2	Mt. Vernon, Squalicum at LC	V F
1	4:19.00	May 9	NWC 2A SubDistrict Meet - Day 1	V F
1	4:19.00	May 11	NWC 2A SubDistrict Meet - Day 2	V F
1	4:21.52	May 16	2A Northwest District 1 - Day 1	V F
1	4:21.52	May 18	2A Northwest District 1 - Day 2	V F
2	4:11.29 PR	May 26	2A WIAA State Championships - All Days	2A F

[Customize]

Training Log

5/8 NWC 2A Sub District Meet	0.99 mi.
5/10 NWC 2A Sub Distirct Meet	2.98 mi.
5/15 2A District 1 & 2 Bi-District Cham…	0.99 mi.
5/17 2A District 1 & 2 Bi-District Cham…	2.98 mi.
5/25 2A, 3A, 4A State Championship …	2.98 mi.

Is this your athlete profile?
Claim this athlete to start tracking your workouts!

[Claim this Athlete]

🏆 **Season Records**

400 Meters - Relay Split

2011 Outdoor	10	55.44c
2012 Outdoor	11	54.84c PR *

800 Meters

2010 Outdoor	9	2:08.53
2011 Outdoor	10	2:03.6h *
2012 Outdoor	11	1:59.3h *
2013 Outdoor	12	1:58.13 PR *

1600 Meters

2010 Outdoor	9	4:29.26
2011 Outdoor	10	4:20.62 *
2012 Outdoor	11	4:11.29 PR *
2013 Outdoor	12	4:11.85

1 Mile

2012 Indoor	11	4:18.23 PR

3000 Meters

Methodology

Athletes:
    400 Meters   - 826 athletes
    1600 Meters - 2056 athletes
    Competed all 4 years in high school
    Graduated between 2006 and 2017

Features
Athlete Name (& ID)
Athlete Sex
9th grade PR
10th grade PR
11th grade PR
12th grade PR
~~School~~
Athletic District
Graduation Year

Athletic Districts in Washington State

Results

Algebraic Model

Assume constant growth rate

$$\frac{t_{12}}{t_{11}} = \frac{t_{11}}{t_{10}}$$
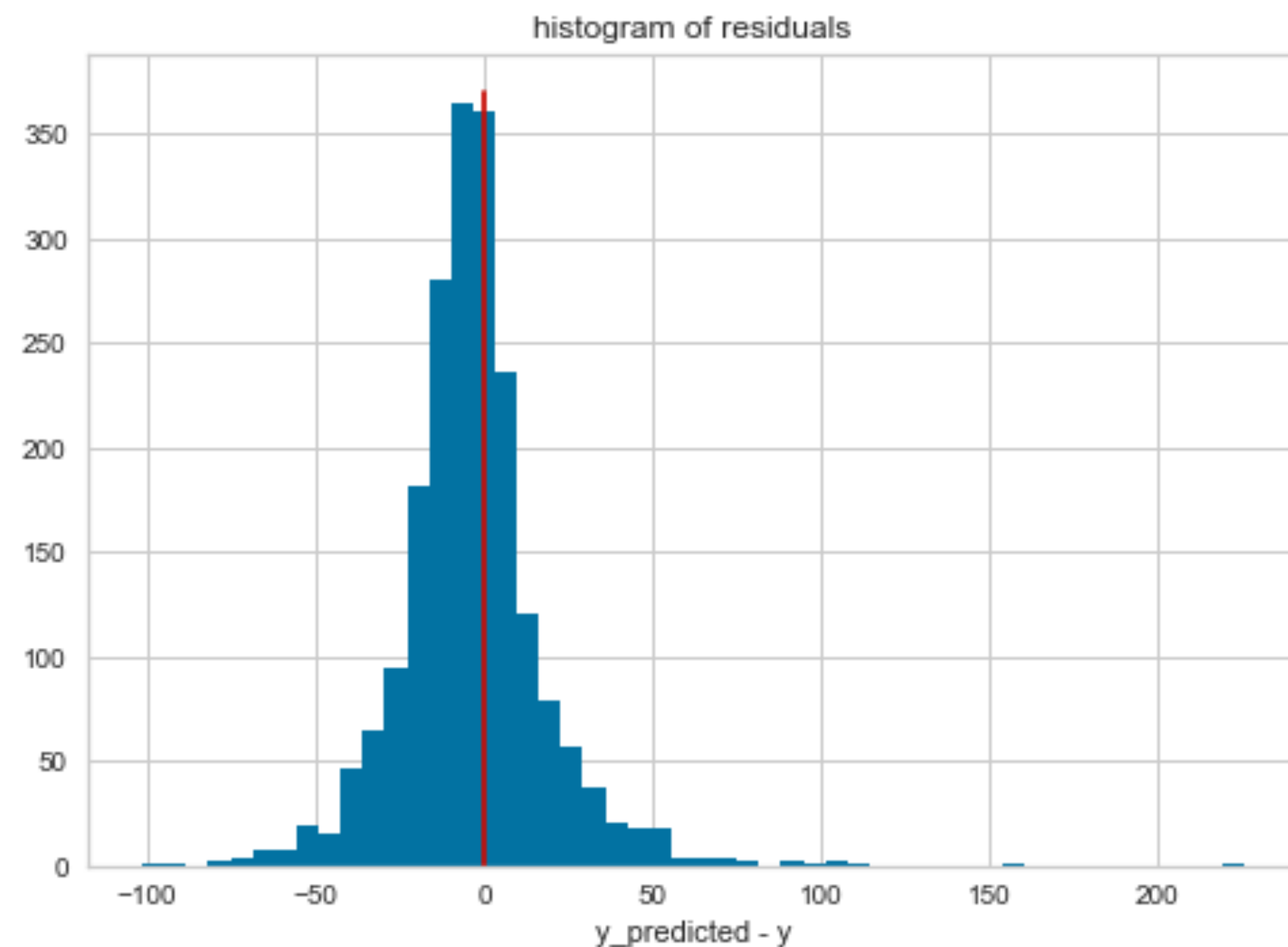
histogram of residuals

400 Meters.   mean = 58.4 seconds

Model	RMSE
Algebra	3.7 (6.3%)
Linear regression	
Mixed Effects Model	
Add higher order terms to LR	

1600 Meters   mean = 317.0 seconds

Model	RMSE
Algebra	21.6 (6.8%)
Linear regression	
Mixed Effects Model	
Add higher order terms to LR	

Linear Regression
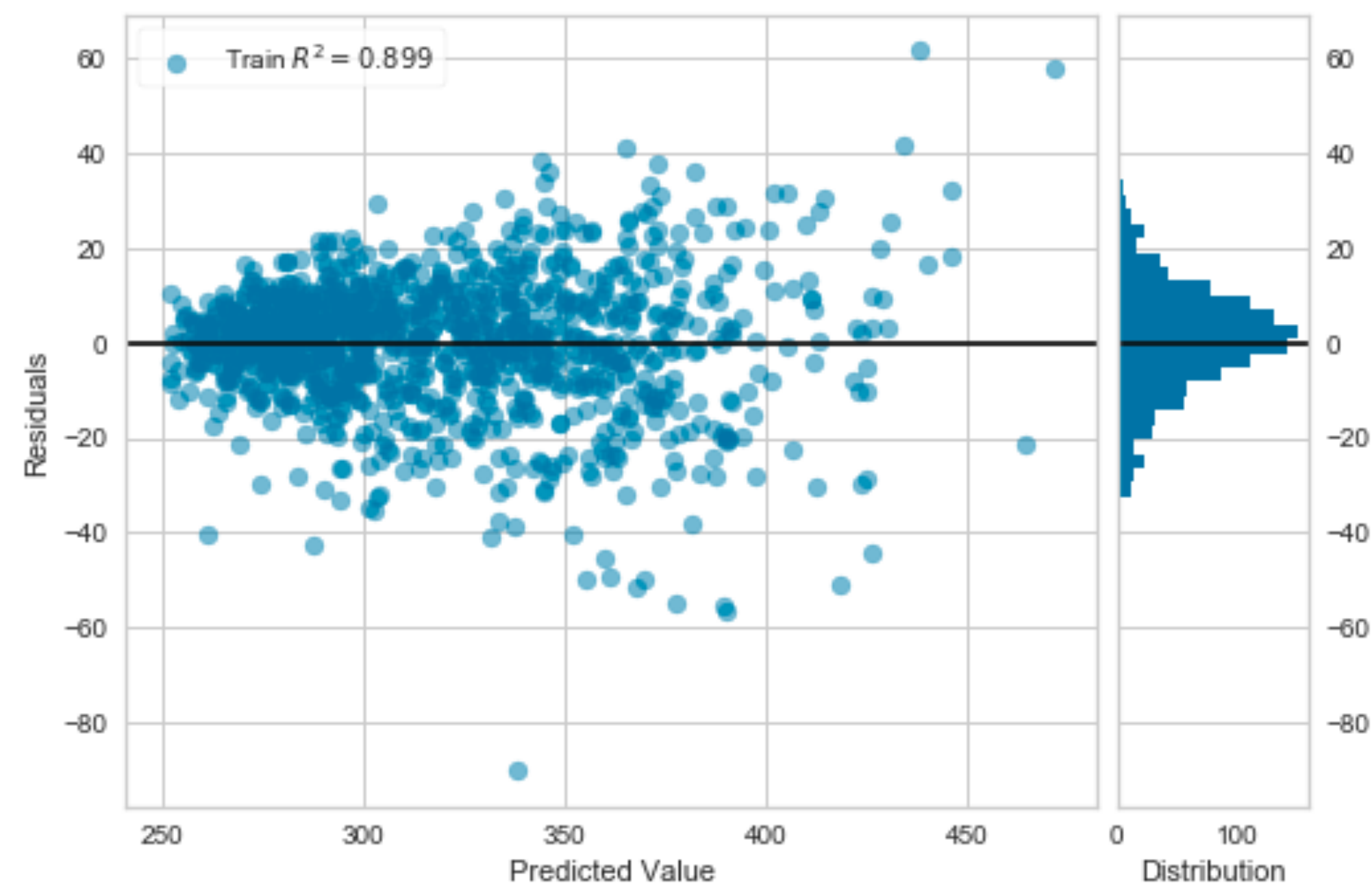
400 Meters.　　　mean = 58.4 seconds

Model	RMSE
Algebra	3.7　(6.3%)
Linear regression	2.2　(3.7%)
Mixed Effects Model	2.1
Add higher order terms to LR	2.3

Mixed Effects Model
　District as mixed effect

Add second order terms for
　$(10^{th}$ grade PR$)^2$
　$(11^{th}$ grade PR$)^2$

1600 Meters　　mean = 317.0 seconds

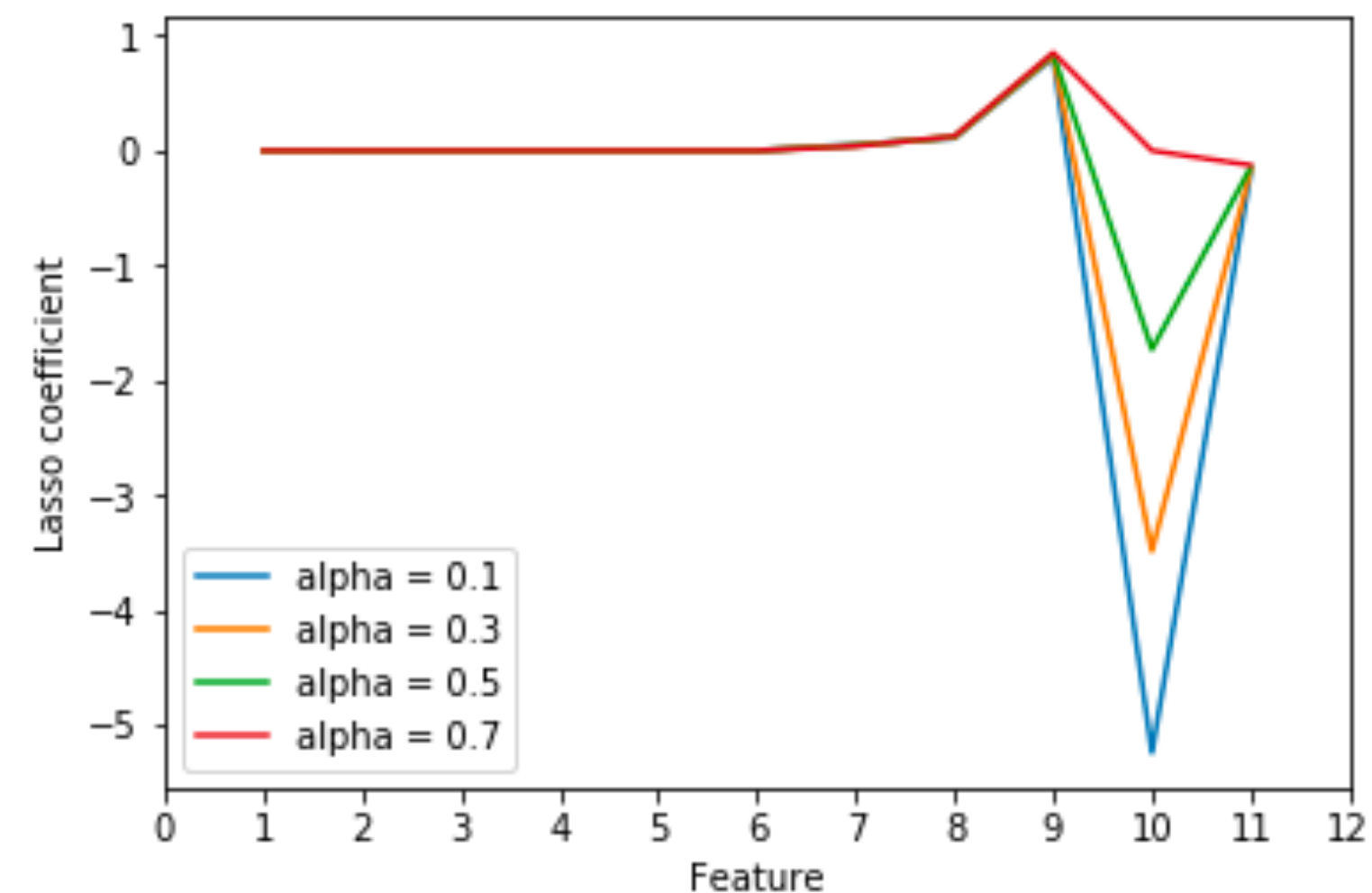| Model	RMSE
Algebra	21.6　(6.8%)
Linear regression	13.5　(4.3%)
Mixed Effects Model	13.6
Add higher order terms to LR	13.4

Lasso
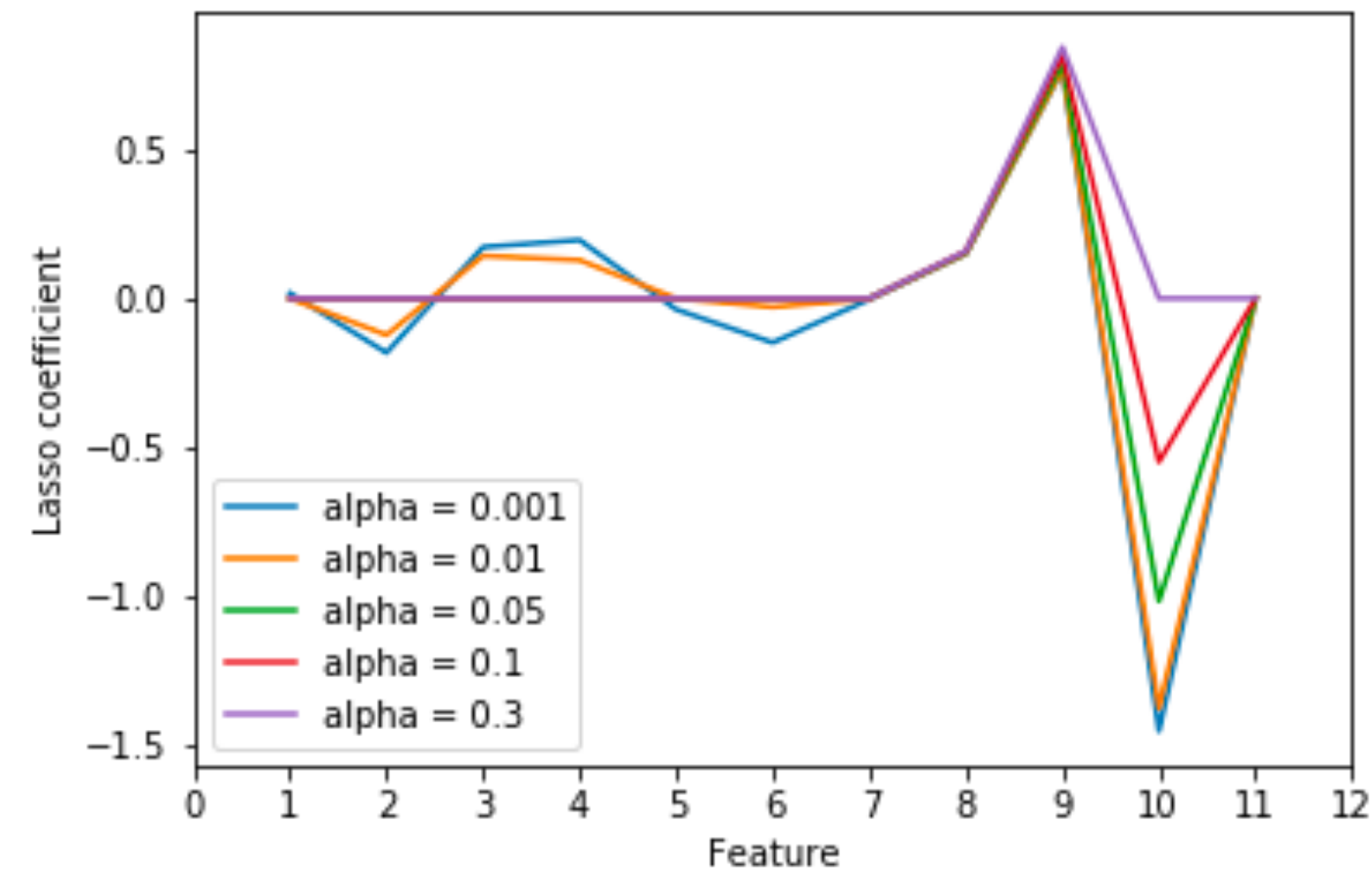
400 Meters

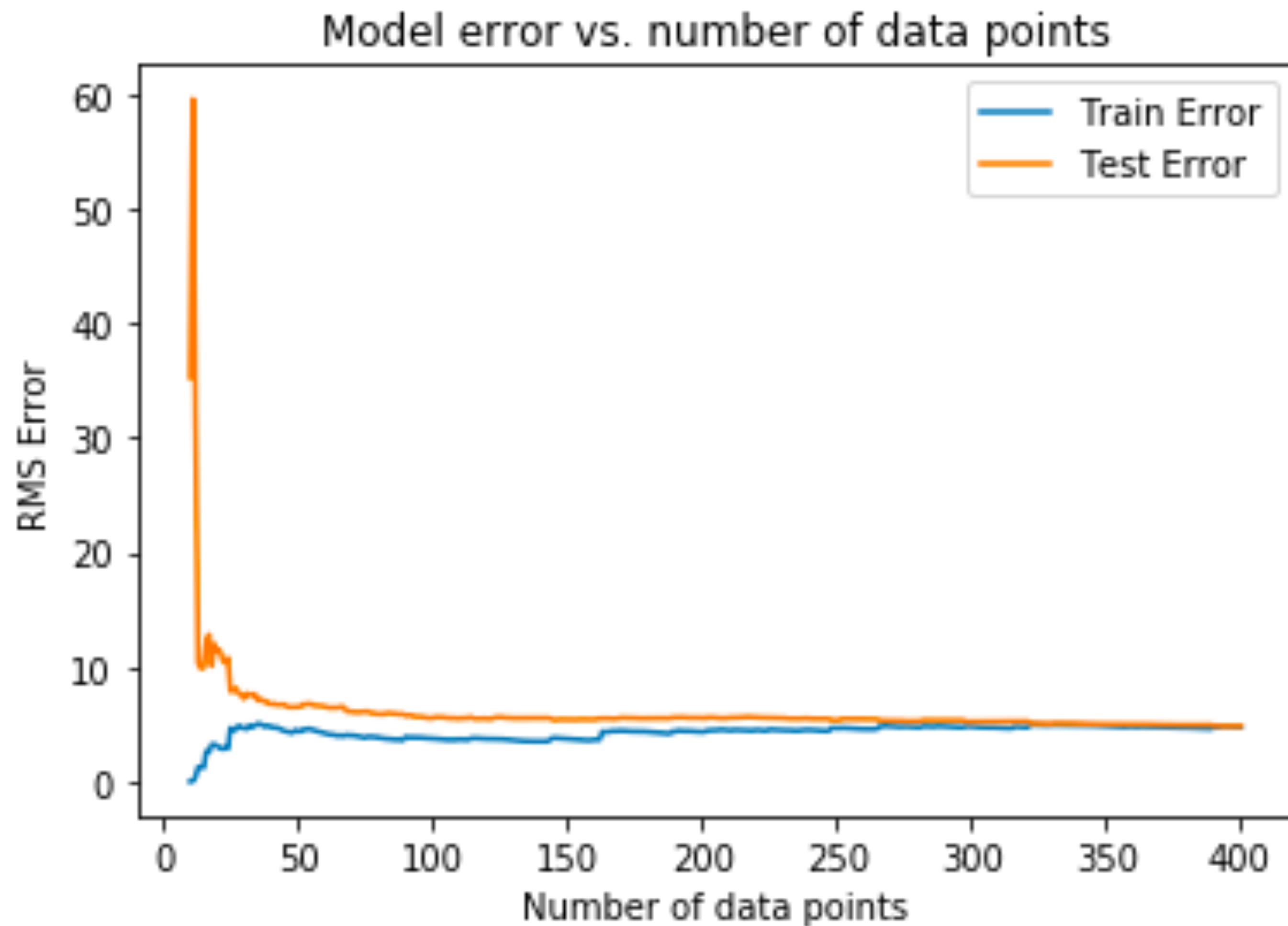- **11th grade time**
- **10th grade time**
- Sex
- Athletic district becomes important

1600 Meters

- **11th grade time**
- Sex
- Graduation year becomes important

Effect of #data points on model

Model error vs. number of data points

It is possible to model with relatively few points

Conclusions

- Machine learning models offer improvement over the algebraic model

- Require more features to get better prediction

Future Work

- Add more features
    - Times for individual races

- Use this technique to predict field events

- Extrapolate to predict college times based on high school times (for recruiting)

Appendix

Data sources:

Washington State Athletic Districts
http://www.wiaa.com/subcontent.aspx?SecID=379

Track & Field data
www.athletic.net