

# Project Luther Proposal

## Predicting High School Track Results

Dana Lindquist  
October 3, 2018

Athletic.net is a non-profit website which compiles Track & Field and Cross Country statistics on athletes from middle school, high school, college and club teams. Each meet's results are recorded and it is possible to see the progression over the years of an individual athlete.

I will focus on high school girls competing in the 400 Meters race. Based on data from 9th, 10th and 11th grade I would like to predict the best time for this athlete in the 12th grade.

There are hundreds of high school athletes with profiles in the athletic.net web site. I plan to focus on athletes who fall into the following categories:

1. Washington State District 1 & District 2 3A schools
2. Athletes who went to the District competition (to filter out athletes who were not putting in a serious effort)
3. Data exists from 2006 to 2018. To make sure I can capture a full 4 years of data for each athlete I will focus on athletes who graduated between 2010 and 2018.

This should give me approximately 150 athletes. I can expand to more districts if more data is required.

Below is an example of data for an individual athlete in one track season.

### 400 Meters

1	1:01.63	Apr 5	Shorecrest, Kamiak, Shoreline Christian, Sitka HS (Alaska)	V F
9	1:00.33	Apr 14	Pasco Invitational	V F
7	1:01.49	Apr 28	Shoreline Invitational (COED)	V F
1	59.75	May 9	WESCO 3A South - Day 1	V P
1	59.75	May 11	WESCO 3A South - Day 2	V P
1	59.17	May 11	WESCO 3A South - Day 2	V F
2	1:01.41	May 16	3A/4A NW District 1 Championships - Day 1	1 P
2	1:01.41	May 18	3A/4A NW District 1 Championships - Day 2	1 P
1	58.95 PR	May 18	3A/4A NW District 1 Championships - Day 2	1 F
12	1:01.01	May 26	WIAA 2A, 3A, 4A State Championship Meet (HS)	3A P

The data I will collect is in the following table. Each athlete will have 11 features.

Variable	Type	Description	Use in Model
Athlete's Name	str	Record descriptor	
Athlete's ID number	int	Record descriptor from athletic.net	
Athlete's School	Int	Each school will be given an integer number	Y
Athlete 9th Grade:			
Best time	float	Best competition time for this school year	Y
Ratio of best to worst time	float	Might be replaced by % improvement from worst to best times	Y
Number of competitions	int	Number of competitions entered in this event during the season	Y
Athlete 10th Grade:			
Best time	float		Y
Ratio of best to worst time	float		Y
Number of competitions	int		Y
Athlete 11th Grade:			
Best time	float		Y
Ratio of best to worst time	float		Y
Number of competitions	int		Y
Athlete 12th Grade:			
Best time	float		Target
Ratio of best to worst time	float		N
Number of competitions	int		N
Number of other events athlete competed in	int	Athletes participate in more than just one event. The number of total events the athlete competed in.	Y

Potential extensions of the project:

- Extend work to boys running the 400 Meters or other events.
- Add data on the schools as well such as average gpa, number of athletes who went to the District competition.