

# Ensemble Models Success Stories

## Lesson 5 – Section 4

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON



## Overview

Examples of How Ensemble Models Helped Improve the Machine Learning Solutions

- Netflix Prize
- KDD Cup 2011

Drawbacks of Ensemble Models

- Cons
- Risk factors

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

# Netflix Prize

## Netflix Prize

*Began October 2006*

### Supervised learning task

- Training data is a set of users and ratings (1,2,3,4,5 stars) those users have given to movies.
- Construct a classifier that given a user and an unrated movie, correctly classifies that movie as either 1, 2, 3, 4, or 5 stars

\$1 million prize for a 10% improvement over Netflix's current movie recommender/classifier  
(MSE = 0.9514)

<http://www.wired.com/business/2009/09/how-the-netflix-prize-was-won/>, a light read (highly suggested)

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

Netflix Prize

meRulesLeaderboardRegisterUpdateSubmitDownload

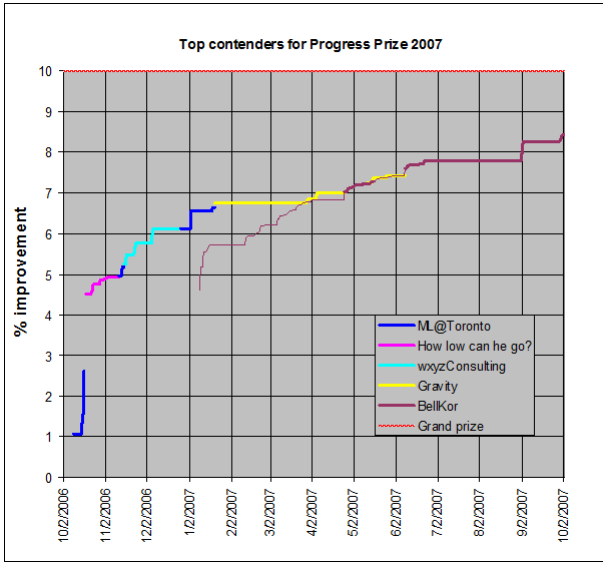
Leaderboard

Team Name	Best Score	% Improvement
No Grand Prize candidates yet		
Grand Prize - RMSE <= 0.8563		
How low can he go?	0.9046	4.92
<a href="#">ML@UToronto A</a>	0.9046	4.92
<a href="#">ssorkin</a>	0.9089	4.47
<a href="#">wxyzconsulting.com</a>	0.9103	4.32
The Thought Gang	0.9113	4.21
<a href="#">NIPS Reject</a>	0.9118	4.16
<a href="#">simonfunk</a>	0.9145	3.88
Bozo_The_Clown	0.9177	3.54
Elliptic Chaos	0.9179	3.52
datcracker	0.9183	3.48
<a href="#">Foreseer</a>	0.9214	3.15
bsdfish	0.9229	3.00
Three Blind Mice	0.9234	2.94
<a href="#">Bocsimacko</a>	0.9238	2.90
Remco	0.9252	2.75
<a href="#">karmatics</a>	0.9301	2.24
Chapelator	0.9314	2.10
<a href="#">Flmod</a>	0.9325	1.99
mthrox	0.9328	1.96

Just three weeks after it began, at least 40 teams had bested the Netflix classifier.

Top teams showed about 5% improvement.

However, improvement slowed...



from <http://www.research.att.com/~volinsky/netflix/>

No Progress Prize candidates yet			
Progress Prize - RMSE <= 0.8625			
1	BelKor	0.8705	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell			
2	KorBell	0.8712	8.43
3	When Gravity and Dinosaurs Unite	0.8717	8.38
4	Gravity	0.8743	8.10
5	basho	0.8746	8.07
6	Dinosaur Planet	0.8753	8.00
7	ML@UToronto A	0.8787	7.64
8	Arek Paterek	0.8789	7.62
9	NIPS Reject	0.8808	7.42
10	Just a guy in a garage	0.8834	7.15
11	Ensemble Experts	0.8841	7.07
12	mathematical capital	0.8844	7.04
13	HowLowCanHeGo2	0.8847	7.01
14	The Thought Gang	0.8849	6.99
15	Reel Ingenuity	0.8855	6.93
16	strudeltamale	0.8859	6.88
17	NIPS Submission	0.8861	6.86
18	Three Blind Mice	0.8869	6.78
19	TrainOnTest	0.8869	6.78
20	Geoff Dean	0.8869	6.78
21	Rookies	0.8872	6.75
22	Paul Harrison	0.8872	6.75
23	ATTEAM	0.8873	6.74
24	wyzconsulting.com	0.8874	6.73
25	ICMLsubmission	0.8875	6.72
26	Efratko	0.8877	6.70
27	Kitty	0.8881	6.65
28	SecondaryResults	0.8884	6.62
29	Birgit Kraft	0.8885	6.61

The top team posted a 8.5% improvement.

Ensemble methods are the best performers...

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

### Rookies

“Thanks to Paul Harrison's collaboration, a simple mix of our solutions improved our result from 6.31 to 6.75”

No Progress Prize candidates yet			
Progress Prize - RMSE <= 0.8625			
1	BelKor	0.8705	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell			
2	KorBell	0.8712	8.43
3	When Gravity and Dinosaurs Unite	0.8717	8.38
4	Gravity	0.8743	8.10
5	basho	0.8746	8.07
6	Dinosaur Planet	0.8753	8.00
7	ML@UToronto A	0.8787	7.64
8	Arek Paterek	0.8789	7.62
9	NIPS Reject	0.8808	7.42
10	Just a guy in a garage	0.8834	7.15
11	Ensemble Experts	0.8841	7.07
12	mathematical capital	0.8844	7.04
13	HowLowCanHeGo2	0.8847	7.01
14	The Thought Gang	0.8849	6.99
15	Reel Ingenuity	0.8855	6.93
16	strudeltamale	0.8859	6.88
17	NIPS Submission	0.8861	6.86
18	Three Blind Mice	0.8869	6.78
19	TrainOnTest	0.8869	6.78
20	Geoff Dean	0.8869	6.78
21	Rookies	0.8872	6.75
22	Paul Harrison	0.8872	6.75
23	ATTEAM	0.8873	6.74
24	wyzconsulting.com	0.8874	6.73
25	ICMLsubmission	0.8875	6.72
26	Efratko	0.8877	6.70
27	Kitty	0.8881	6.65
28	SecondaryResults	0.8884	6.62
29	Birgit Kraft	0.8885	6.61

CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Arek Paterek

“My approach is to **combine the results of many methods** (also two-way interactions between them) using linear regression on the test set. The best method in my ensemble is regularized SVD with biases, post processed with kernel ridge regression”

[http://rainbow.mimuw.edu.pl/~ap/ap\\_kdd.pdf](http://rainbow.mimuw.edu.pl/~ap/ap_kdd.pdf)

No Progress Prize candidates yet			
Progress Prize - RMSE <= 0.8625			
1	BellKor	0.8705	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell			
2	KorBell	0.8712	8.43
3	When Gravity and Dinosaurs Unite	0.8717	8.38
4	Gravity	0.8743	8.10
5	basho	0.8746	8.07
6	Dinosaur Planet	0.8753	8.00
7	ML@UToronto A	0.8787	7.64
8	Arek Paterek	0.8789	7.62
9	NIPS Reject	0.8808	7.42
10	Just a guy in a garage	0.8834	7.15
11	Ensemble Experts	0.8841	7.07
12	mathematical capital	0.8844	7.04
13	HowLowCanHeGo2	0.8847	7.01
14	The Thought Gang	0.8849	6.99
15	Reel Ingenuity	0.8855	6.93
16	strudeltamale	0.8859	6.88
17	NIPS Submission	0.8861	6.86
18	Three Blind Mice	0.8869	6.78
19	TrainOnTest	0.8869	6.78
20	Geoff Dean	0.8869	6.78
21	Rookies	0.8872	6.75
22	Paul Harrison	0.8872	6.75
23	ATTEAM	0.8873	6.74
24	wyzconsulting.com	0.8874	6.73
25	ICMLsubmission	0.8875	6.72
26	Efratko	0.8877	6.70
27	Kitty	0.8881	6.65
28	SecondaryResults	0.8884	6.62
29	Birgit Kraft	0.8885	6.61

CONTINUING EDUCATION  
UNIVERSITY OF WASHINGTON

## U of Toronto

“When the predictions of **multiple** RBM models and **multiple** SVD models are linearly combined, we achieve an error rate that is well over 6% better than the score of Netflix’s own system.”

<http://www.cs.toronto.edu/~rsalakhu/papers/rbmc.pdf>

No Progress Prize candidates yet			
Progress Prize - RMSE <= 0.8625			
1	BellKor	0.8705	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell			
2	KorBell	0.8712	8.43
3	When Gravity and Dinosaurs Unite	0.8717	8.38
4	Gravity	0.8743	8.10
5	basho	0.8746	8.07
6	Dinosaur Planet	0.8753	8.00
7	ML@UToronto A	0.8787	7.64
8	Arek Paterek	0.8789	7.62
9	NIPS Reject	0.8808	7.42
10	Just a guy in a garage	0.8834	7.15
11	Ensemble Experts	0.8841	7.07
12	mathematical capital	0.8844	7.04
13	HowLowCanHeGo2	0.8847	7.01
14	The Thought Gang	0.8849	6.99
15	Reel Ingenuity	0.8855	6.93
16	strudeltamale	0.8859	6.88
17	NIPS Submission	0.8861	6.86
18	Three Blind Mice	0.8869	6.78
19	TrainOnTest	0.8869	6.78
20	Geoff Dean	0.8869	6.78
21	Rookies	0.8872	6.75
22	Paul Harrison	0.8872	6.75
23	ATTEAM	0.8873	6.74
24	wyzconsulting.com	0.8874	6.73
25	ICMLsubmission	0.8875	6.72
26	Efratko	0.8877	6.70
27	Kitty	0.8881	6.65
28	SecondaryResults	0.8884	6.62
29	Birgit Kraft	0.8885	6.61

CONTINUING EDUCATION  
UNIVERSITY OF WASHINGTON

# Gravity

Table 5: Best results of single approaches and their combinations

Method/Combination	RMSE
MF	0.9190
NB	0.9313
CL	0.9606
NB + CL	0.9275
MF + CL	0.9137
MF + NB	0.9089
MF + NB + CL	0.9089

home.mit.bme.hu/~gtakacs/download/gravity.pdf

No Progress Prize candidates yet		
Progress Prize - RMSE <= 0.8625		
1	BellKor	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell		
2	KorBell	8.43
3	When Gravity and Dinosaurs Unite	8.38
4	Gravity	8.10
5	basho	8.07
6	Dinosaur Planet	8.00
7	ML@UToronto A	7.64
8	Arek Paterek	7.62
9	NIPS Reject	7.42
10	Just a guy in a garage	7.15
11	Ensemble Experts	7.07
12	mathematical capital	7.04
13	HowLowCanHeGo2	7.01
14	The Thought Gang	6.99
15	Reel Ingenuity	6.93
16	strudeltamale	6.88
17	NIPS Submission	6.86
18	Three Blind Mice	6.78
19	TrainOnTest	6.78
20	Geoff Dean	6.78
21	Rookies	6.75
22	Paul Harrison	6.75
23	ATTEAM	6.74
24	wyzconsulting.com	6.73
25	ICMLsubmission	6.72
26	Efratko	6.70
27	Kitty	6.65
28	SecondaryResults	6.62
29	Birgit Kraft	6.61

CONTINUING EDUCATION  
UNIVERSITY OF WASHINGTON

# When Gravity and Dinosaurs Unite

“Our common team blends the result of team Gravity and team Dinosaur Planet.”

Might have guessed from the name...

No Progress Prize candidates yet		
Progress Prize - RMSE <= 0.8625		
1	BellKor	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell		
2	KorBell	8.43
3	When Gravity and Dinosaurs Unite	8.38
4	Gravity	8.10
5	basho	8.07
6	Dinosaur Planet	8.00
7	ML@UToronto A	7.64
8	Arek Paterek	7.62
9	NIPS Reject	7.42
10	Just a guy in a garage	7.15
11	Ensemble Experts	7.07
12	mathematical capital	7.04
13	HowLowCanHeGo2	7.01
14	The Thought Gang	6.99
15	Reel Ingenuity	6.93
16	strudeltamale	6.88
17	NIPS Submission	6.86
18	Three Blind Mice	6.78
19	TrainOnTest	6.78
20	Geoff Dean	6.78
21	Rookies	6.75
22	Paul Harrison	6.75
23	ATTEAM	6.74
24	wyzconsulting.com	6.73
25	ICMLsubmission	6.72
26	Efratko	6.70
27	Kitty	6.65
28	SecondaryResults	6.62
29	Birgit Kraft	6.61

CONTINUING EDUCATION  
UNIVERSITY OF WASHINGTON

# BellKor / KorBell

And, yes, the top team which is from AT&T...

“Our final solution (RMSE=0.8712) consists of blending 107 individual results.”

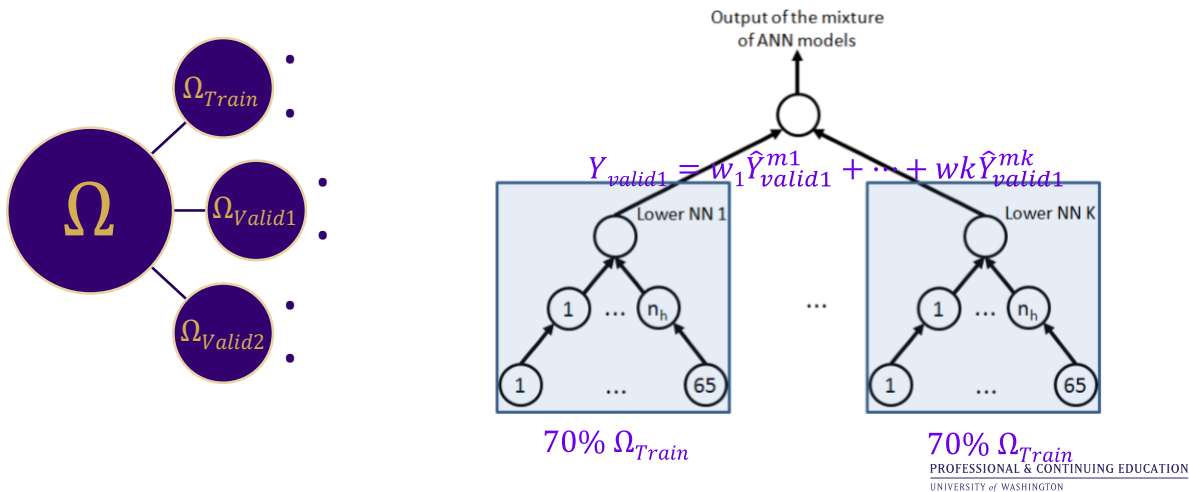
No Progress Prize candidates yet			
Progress Prize - RMSE <= 0.8625			
1	BellKor	0.8705	8.50
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell			
2	KorBell	0.8712	8.43
3	When Gravity and Dinosaurs Unite	0.8717	8.38
4	Gravity	0.8743	8.10
5	basho	0.8746	8.07
6	Dinosaur Planet	0.8753	8.00
7	ML@UToronto A	0.8787	7.64
8	Arek Paterek	0.8789	7.62
9	NIPS Reject	0.8808	7.42
10	Just a guy in a garage	0.8834	7.15
11	Ensemble Experts	0.8841	7.07
12	mathematical capital	0.8844	7.04
13	HowLowCanHeGo2	0.8847	7.01
14	The Thought Gang	0.8849	6.99
15	Reel Ingenuity	0.8855	6.93
16	strudeltamale	0.8859	6.88
17	NIPS Submission	0.8861	6.86
18	Three Blind Mice	0.8869	6.78
19	TrainOnTest	0.8869	6.78
20	Geoff Dean	0.8869	6.78
21	Rookies	0.8872	6.75
22	Paul Harrison	0.8872	6.75
23	ATTEAM	0.8873	6.74
24	wyzconsulting.com	0.8874	6.73
25	ICMLsubmission	0.8875	6.72
26	Efratko	0.8877	6.70
27	Kitty	0.8881	6.65
28	SecondaryResults	0.8884	6.62
29	Birgit Kraft	0.8885	6.61

CONTINUING EDUCATION  
UNIVERSITY OF WASHINGTON

# KDD Cup Example

# KDD Cup 2011

- Predict whether a user is going to rate a music track highly or not



## Performance Gain Decreases as Ensembled Models Increase

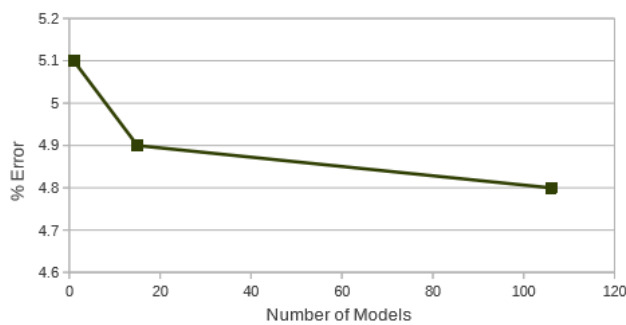


Figure 5: Error on Combinations of Neural Nets.



## Cons of Ensemble of Models

- Loss of interpretability quickly
- Difficult to operationalize
- Risk of overfitting
  - Heritage Health Prize: prediction number of hospital stay days of next year
  - Ensemble of around 50 models
  - Ranked #1 on public leaderboard, but ranked #20 on private leaderboard

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Risk Factors of Overfitting in Ensemble Models

- Highly correlated models:
  - Different models trained on same or similar feature set
  - Similar models trained on similar feature sets
- Weights of models are fine tuned too much in order to optimize the performance on validation data
  - Overfitting on validation data
  - Does not generalize well in production

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Lesson 5 Summary

- >Random Forest and Gradient Boosted Decision Trees
  - AdaBoost
- >Practiced in Python
  - sklearn.ensemble
- >Random Forest and Gradient Boosted Decision Trees are powerful regression and classification algorithms.

