What makes a good muffin?

Ivan Ivanov
CS 229 Project

Banana Muffins II Recipe - Allrecipes.com
allrecipes.com/recipe/banana-muffins-ii/
Rating: 4.8 - 2,379 reviews - 35 min - 187 cal
These delicious banana muffins are easy for kids to make.

Banana Muffins Recipe : Food Network Kitchen : Food Network
www.foodnetwork.com/recipe/banana-muffins
Rating: 3.8 - 77 reviews - 45 min
Preheat the oven to 425 degrees F. Lightly brush a 12-muffin tin with butter and set aside. Whisk the flour, baking soda, brown sugar, and cinnamon together in a medium bowl; set aside. Whisk the banana, oil, milk, eggs, salt, and vanilla in a large measuring cup with a spout or ...
Motivation and project aims

The problem: finding a good recipe is hard

– is the best rated recipe really the most delicious?
– what if I want to use soy milk instead?
– well, I don’t have any buttermilk...
– how should I split three eggs in half?!

Project aim: develop algorithm to suggest optimal muffin recipe

– predict recipe success based on ingredients
– allow for substitutions
– suggest scaling relationships
Data collection and feature refinement

Recipe:
id: 228553,
link: `http://allrecipes.com/…`
madeitCount: 274,
name: `Moist Chocolate Muffins`,
rating: 4.43,
reviewCount: 183,
servings: 12

Ingredients:
id: 1684,
name: `flour`,
modifiers: `all-purpose`,
amount: 2
unit: `cup`,
…}
Predicting success of banana muffins

Logistic regression successfully predicts recipe outcome

PCA analysis suggests some data structure
Regression model of muffin success

Regression model would allow for optimizing ingredients to maximize recipe success
Predicting success of any muffin

Data on all muffins displays complex structure. Linear regression is not reliable in predicting recipe success. Model suffers from high bias and high variance.
SVM Classification with Gaussian kernel does not reliably predict recipe outcome.

Conclusions

– Logistic regression successfully predicts banana muffin recipe success.
– Generalizing the model to a wider variety of muffin recipes presents a challenge, which may be solved by collecting more data.
– Successful development of this software will remove guesswork from cooking to deliver robust results.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted:</th>
<th>N = 108</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>Bad</td>
<td>22</td>
<td>42</td>
</tr>
<tr>
<td>Good</td>
<td>6</td>
<td>38</td>
</tr>
</tbody>
</table>