Online reviews of smartphone applications often provide a neutral overall rating using the Likert or semantic scales but these do not fully reveal sentiments of customers about a product’s features. Google Play captures both structured and unstructured reviews generic to each application. This leaves the user with no clue about a particular feature and its rating. In this paper, we illustrate feature extraction and feature ranking of the Photosharing application using SAS® Sentiment Studio. Sentiment Studio is used to predict a new review as either Positive or Non-positive. We built a model rules in the Rule-Based model. In this data, users distinguish between the statistical model or the hybrid model in a test data set. The best model helps us in categorizing each review of the application by its feature along with the rating.

**Figure 1: Training results for the Statistical Model**

![Graphical results for the Statistical Model](image)

**TEST FOR THE STATISTICAL MODEL**

A stratified sample of 50 documents was considered for testing the models. The Precision in scoring came out about 90%. Below are the screenshots of a few scored documents.

**Documents correctly predicted as Non-Positive:**

<table>
<thead>
<tr>
<th>ID</th>
<th>Comment</th>
<th>Rating</th>
<th>Positive</th>
<th>Negative</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“It’s a great app”</td>
<td>5.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>“The interface is user-friendly”</td>
<td>4.5</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Documents correctly predicted as Positive:**

<table>
<thead>
<tr>
<th>ID</th>
<th>Comment</th>
<th>Rating</th>
<th>Positive</th>
<th>Negative</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>“It’s a great app”</td>
<td>5.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>“The interface is user-friendly”</td>
<td>4.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Figure 2: Graphical results for the Statistical Model**

![Graphical results for the Statistical Model](image)

**TEST FOR RULE-BASED MODEL**

The Rule-based model is more sophisticated compared to the Statistical model. This model allows us to write custom rules along with the rules learned from the Statistical model. The Rule-Based model provides facility for us to define products and features specific to our project to identify sentiment at the granularity level or feature level. The rules imported into Tonal Keywords are classified as “Classified” by default.

**RULE-BASED MODEL BUILDING**

The Rule-based model is more sophisticated compared to the Statistical model. This model allows us to write custom rules along with the rules learned from the Statistical model. The Rule-Based model provides facility for us to define products and features specific to our project to identify sentiment at the granularity level or feature level. The rules imported into Tonal Keywords are classified as “Classified” by default.

**FEATURES IDENTIFICATION**

We have defined Photosharp as a Product: Editing, Photo史料, Updating and General are the features for the product Photosharp. Different terms that identify the features are defined in the definitions portion of each feature.

<table>
<thead>
<tr>
<th>Features</th>
<th>Attributes that capture the definition of a feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editing</td>
<td>Editing, add, filter, sharpen, toning, picture-quality, picture-tool</td>
</tr>
<tr>
<td>Photoshare</td>
<td>Picture size, photo size, crop size, square, large, small</td>
</tr>
<tr>
<td>Updating</td>
<td>Update, upload, export, enhance, close, fix, share, print</td>
</tr>
<tr>
<td>General</td>
<td>Follow, beautiful, keep-up, simple, add, edit, great, amazing, fix, easy</td>
</tr>
</tbody>
</table>

**Figure 3: Test results of the Statistical Model for a document predicted**

![Graphical results for the Statistical Model](image)

**Figure 4: Test results of the Statistical Model for a document predicted**

![Graphical results for the Statistical Model](image)

**Figure 5: Overall results of the Statistical Model on Testing Documents**

![Graphical results for the Statistical Model](image)

**Statistical model test results on the scoring data with 19 Positive comments out of 24 are correctly predicted with an accuracy being 79% and 26 Non-positive comments out of 26 are correctly predicted with an accuracy of 100%. However the model has 20.8% of false negative rate (Type II error).**

**Figure 6: Overall results of Rule Based Model on Testing Documents**

![Graphical results for the Statistical Model](image)

**Testing the Rule-Based model has given excellent results with 23 out of 24 Positive comments being correctly predicted with an accuracy of 95% and 26 out of 26 Non-positive comments being correctly predicted with an accuracy of 100%. Rule-Based model has a decreased false negative rate (Type II error) of 0.5%.”**

**HYBRID MODEL**

The Hybrid model is a combination of the Rule-Based and Statistical models by applying the rules developed in Rule-Based and the numerical data from the activated Statistical model.

**Figure 7: Feature level sentiment detection using Hybrid model**

![Graphical results for the Statistical Model](image)

**Figure 8: Overall results of the Hybrid model on Testing Documents**

![Graphical results for the Statistical Model](image)

**MODEL COMPARISON**

<table>
<thead>
<tr>
<th>Model</th>
<th>Positive Precision</th>
<th>Negative Precision</th>
<th>False Negative Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>100%</td>
<td>100%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Rule-Based</td>
<td>95%</td>
<td>100%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Hybrid</td>
<td>100%</td>
<td>100%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

**References**


**Acknowledgments**

The authors would like to thank Dr. Goutam Chakraborty for his invaluable supervision and support throughout this research.

**Contact Information**

Your comments and questions are valued and encouraged. Contact the author at:

Name: Siddhartha Reddy Mandati  
Company: Oklahoma State University  
E-mail: siddhartha.mandati@gmail.com

Name: Anil Kumar Pantangi  
Company: Oklahoma State University  
E-mail: anil.pantangi@okstate.edu

Name: Goutam Chakraborty  
Company: Oklahoma State University  
E-mail: goutam.chakraborty@okstate.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.