Training/Testing Split:

80% / 20%

Text Vectorization Methods:

1. Countvectorizer
2. TfidfVectorizer

Classification methods:

1. Loop through labels and do binary classification for each label with Multinomial Naive Bayes
2. OneVsRestClassifier implemented by Multinomial Naive Bayes classifier

Result:

**OneVsRestClassifier**

|  | Count vectorizer | Tfidf vectorizer |
| --- | --- | --- |
| Score | **0.050082918739635156** | 0.022222222222222223 |

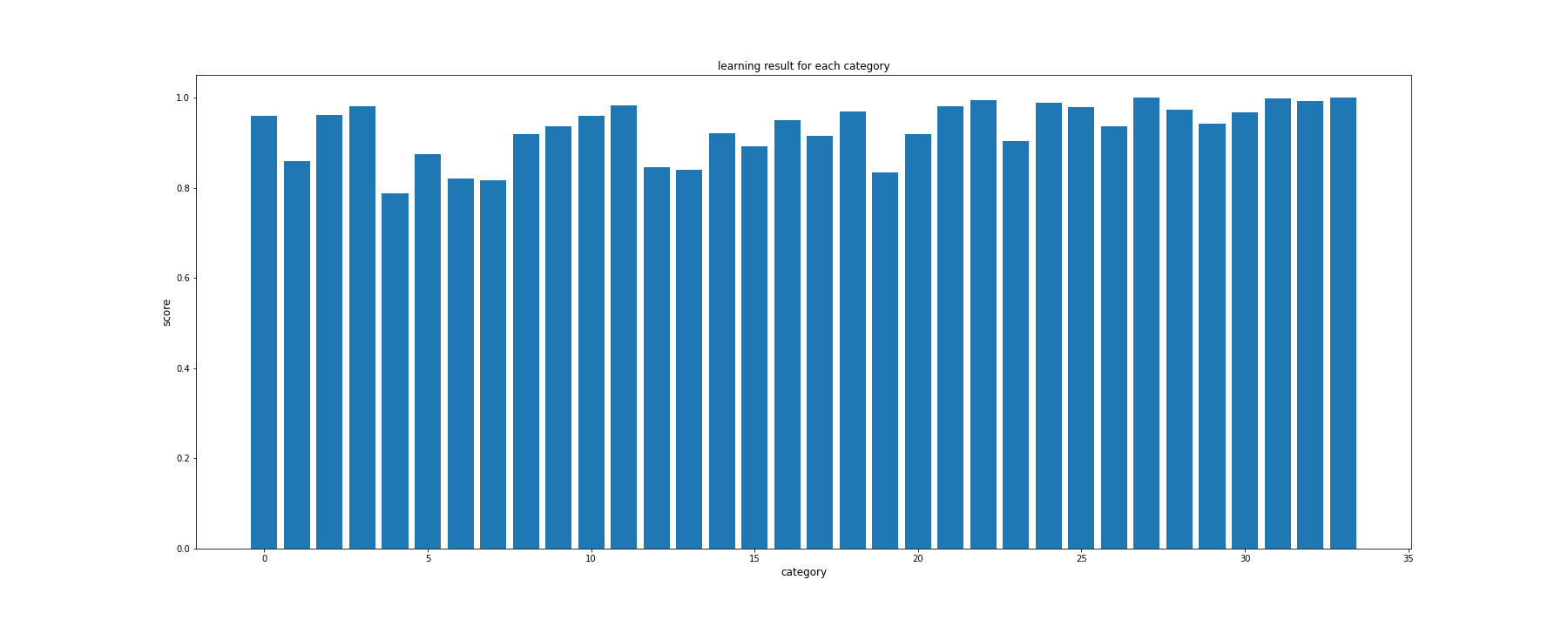
**Multiple Naive Bayes classifier on labels**

Countvectorizer

| Category | MultinomialNB Using Countvectorizer | MultinomialNB Using TfidfVectorizer |
| --- | --- | --- |
| **Animal** | 0.9592039800995025 | **0.9996683250414594** |
| **Partying & Bad Behavior** | 0.8587064676616916 | **0.8620232172470979** |
| **Miscellaneous** | 0.9611940298507463 | **0.9651741293532339** |
| **Blue Collar** | 0.9804311774461029 | **0.9996683250414594** |
| **Men/Women** | 0.7873963515754561 | **0.8199004975124378** |
| **Money** | 0.8739635157545605 | **0.8805970149253731** |
| **Lookin' Good** | 0.820563847429519 | **0.854726368159204** |
| **Dirty** | 0.8172470978441128 | **0.8258706467661692** |
| **Marriage** | 0.9187396351575456 | **0.96318407960199** |
| **Technology** | 0.936318407960199 | **0.9379767827529022** |
| **Doctor** | 0.9588723051409619 | **0.9993366500829187** |
| **Food** | 0.9830845771144279 | **0.9996683250414594** |
| **Nationality** | 0.8454394693200663 | **0.8470978441127694** |
| **News & Politics** | **0.8398009950248756** | 0.83681592039801 |
| **Police & Military** | 0.9217247097844112 | **0.9320066334991708** |
| **Travel & Car** | 0.8925373134328358 | **0.8998341625207297** |
| **Insults** | 0.9499170812603649 | **0.9525704809286899** |
| **Sports & Athletes** | 0.9154228855721394 | **0.9233830845771144** |
| **Gross** | 0.9701492537313433 | **0.9993366500829187** |
| **Pop Culture & Celebrity** | **0.8338308457711443** | 0.8325041459369817 |
| **Kids** | 0.9194029850746268 | **0.9645107794361526** |
| **Yo' Mama** | **0.9807628524046434** | 0.9064676616915422 |
| **Lawyer** | 0.9950248756218906 | **1.0** |
| **Work** | 0.9044776119402985 | **0.9164179104477612** |
| **Blonde** | 0.988391376451078 | **0.9993366500829187** |
| **Dark Humor** | 0.9787728026533997 | **0.9996683250414594** |
| **Pick-Up Lines** | **0.9369817578772802** | 0.9346600331674959 |
| **Cross the Road** | 0.9996683250414594 | **1.0** |
| **God** | 0.9734660033167496 | **0.9996683250414594** |
| **School** | 0.9422885572139303 | **0.9459369817578773** |
| **Fat** | 0.9678275290215589 | **0.9996683250414594** |
| **Little Johnny** | 0.9983416252072969 | **1.0** |
| **Walks into a Bar** | 0.9930348258706467 | **0.9946932006633499** |
| **NSFW** | 0.9996683250414594 | **1.0** |

TfidfVectorizer

Count Vectorized score:



Tfidf Vectorized score:

