AMPEL: An Approach for Machine-learning Based Prediction and Evaluation of the Learned Success of Social Media Posts

Max-Emanuel Keller, Alexander Döschl, Peter Mandl
HM Hochschule München University of Applied Sciences
Munich, Germany
{max-emanuel.keller, alexander.doeschl, peter.mandl}@hm.edu

***Abstract*—In this paper, we present AMPEL, a system that assists social media managers in creating successful posts for company pages on social media platforms. The AMPEL workflow classifies posts as either successful or unsuccessful, using Facebook as an example. The system makes a prediction of success for a new post during its creation, prior to its publication, which is a major advantage in comparison to existing systems. Posts that are classified as unsuccessful can be revised by the author until the prediction is successful. The system also evaluates previously published posts for success, allowing for comparison between the predicted success and actual success achieved. The two classification models are built using Random Forest, XGBoost, and neural network-based classification algorithms. The system is evaluated using two separate corpora of posts from different industries. We also demonstrate a prototype of AMPEL and show that it achieves good results with very reliable predictions of success. Our evaluation shows that AMPEL can replace manual review by a human social media manager in many applications.**

I. Introduction

Social media portals have become a crucial marketing channel for companies, due to their great popularity among users [1]. As a result, many companies maintain social media pages for their brands to interact with customers and promote the company, its products, and services [2]. This communication can take many forms, but always aims to be received positively by the target audience. Many companies tend to launch long-term marketing campaigns on social media, which often consist of multiple posts. This requires a significant amount of time and resources, resulting in high costs. As the success of a campaign can have a significant impact on the company's overall performance, it's important to create posts that are as successful as possible, while avoiding those that are unsuccessful. Achieving this goal requires answering two questions:

- When is a post considered as successful?
- Which parameters influence the success of a post?

A strong indicator for the success of a post is the way readers interact with it [3], [4], by sending different reactions, replying with comments or sharing the post. Successful posts can be recognized by the fact that they are frequently shared, many comments are left and the positive reactions (like, wow, haha) clearly outweigh the bad ones (angry, sad). The individual definition of success is somewhat subjective and can vary between companies.

Obviously, the attributes that determine the success of a post can also differ between companies and industries. Nevertheless, we can assume that companies with similar business conditions and a comparable communication style will be perceived the same way. In addition to obvious characteristics such as the topic and writing style of a post, more hidden features like the time of publication can also have a major impact on user response.

In this paper, we describe an approach that supports the creation of more successful posts by machine learning assisted decisions based on other posts. This effort seeks to gain an understanding of the factors for success and apply them to the creation of future posts.

II. Problem

Creating high-quality content for business pages on social media portals requires a significant investment of time and experience. One major challenge is selecting appropriate topics for communication. This requires not only a deep understanding of the company, but also a broad knowledge of current trends and popular vocabulary and structures, which can vary between industries. It's important to note that what is popular today may not be relevant or effective in the future.

In addition to obvious attributes influencing success, there may be other hidden parameters that affect the success of a post too. The often unknown relationships between these attributes make the creation of successful social media content a complex, error-prone task. For this reason, many companies already use social media analysis tools. However, systems providing information about the potential success of posts that have not been published yet were not researched in detail so far [1].

Even after a post is published, the work is not finished as it needs to be evaluated over its entire life cycle, which leads to a lot of work as the number of posts increases. To ensure timely response to any undesirable trends, regular monitoring is also crucial.

In order to evaluate the success of posts depending on their topics, annotated posts are needed to classify the content of a post by its topic and evaluate its success. Consequently, we created two corpora that consist of posts from Facebook and are written in German language. The first corpus includes posts from industries including automotive, food retail, and IT. Whereas the second corpus is made of posts taken from the food delivery services industry in Germany. The posts belong to six of the most important brands in the industry.

The annotation was done by a group of experts who annotated each post by topic and success. To annotate the topic, the experts created a list of 11 thematic categories of corporate communication. Each post was labeled thematically according to these categories. Success, on the other hand, was labeled as successful or unsuccessful. The corpora offer the possibility to study the success of posts depending on their topic. We aim to develop a system that predicts the success of a post before its publication and evaluates the actual success after its publication.

III. Approach

In this work we propose AMPEL, an **A**pproach for **M**achine-learning based **P**rediction and **E**valuation of the **L**earned success of social media posts.

AMPEL aims to realize a workflow we call *AMPEL assisted post creation*, shown in Fig. 1. This workflow has the goal of simplifying the process of creating posts and monitoring their success. It supports a social media manager in the creation of posts by providing predictions and ratings of the posts success.

The workflow consists of seven steps. First, the manager writes a new post (step 1), for which he receives a prediction of potential success (step 2), that is either *successful* or *unsuccessful*. If the rating is *unsuccessful*, the content of the post can be revised and the prediction repeated until it is *successful*. Once done, the post is published (step 3). After publishing, the post is read by the target audience, which are the users of the page. The users express their opinion by interacting with the post. This includes giving a reaction, creating a comment or sharing the post. One month after publication of the post, when the vast majority of user interactions have statistically already been made [5], [6], the interactions are retrieved (step 4) and the actual success (step 5) can be rated as either *successful* or *not successful*. Finally (step 6), the predicted success of the post from step 2 is compared to the actual success from step 5 to determine if the prediction was met.

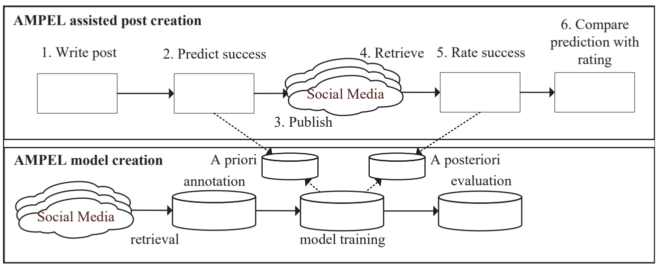

Fig. 1. AMPEL approach

IV. Methodology

AMPEL assisted post creation is based on two classification models, shown in Fig. 1, that can classify the success of a post at two points in time. The first model allows to predict the success of a post even before its publication (step 2), i.e. *a priori*, whereas the second model rates the actual success of a published post retrospectively (step 5), i.e. *a posteriori*.

Both models are created on posts from the history of selected pages that contain the data shown in Table I. The data can be divided into three sources:

- The *original posts*, with fields like date and text,
- their manual *annotations* by human experts, including the success and a category, and
- the *interactions* of the users from Facebook, that contain the number of comments, shares and different types of reactions (likes, haha, wow, etc.).

Both models use *success* as the target label but differ in their features. The features of the *a priori* model include the data of the original post, that are added during the creation process, but also the annotated category. For the *a posteriori* model, the interactions of the users from Facebook that are only present after a post is published are added as additional features.

TABLE I. Data of the models

source	data	a priori	a posteriori
annotation	success category	label feature	label feature
original post	date, text, etc.	feature	feature
interactions	#comments #shares #reactions	-	feature

The models are created during the *AMPEL model creation*, shown in Fig. 1. The process begins with the retrieval of data from the social media platform Facebook using the Graph API. The Graph API is a REST interface that provides various endpoints for writing and retrieving data. For our use case, selected pages are retrieved with a limited history of posts, where the posts also include the corresponding user interactions, which are reactions (like, wow, haha, angry, sad and thankful) and user comments. In a second step, a group of human experts annotates the posts by success and topic.

Since there is no universally accepted definition for the success of a post, this needs to be established. Therefore, we use the knowledge of our social media experts to create a common definition of success that is shared by all annotators and used when annotating success. The resulting data is used to train the a priori and a posteriori models and finally evaluate it.

In the following sections, we describe the individual steps of our approach in greater detail. Our models rely on two existing data sets with annotated posts, which we describe in Section V. In Section VI, the data is used for model creation.

Finally, we evaluate our system in Section VII and give an outlook on a prototype that allows the practical application of the approach in Section VIII.

V. Data sets

The AMPEL approach can be based on arbitrary datasets, assuming the included posts are labeled by their success. Since the thematic range of the company pages on Facebook is enormous, the success factors can also vary widely. To apply the approach presented in this paper to different industries, we create two new corpora.

Both corpora consist of posts published by companies on their pages written in German language. Each post includes an ID, the post text, a publication date, and other media associated with it, such as images, videos, and links. Moreover, each post may include interactions created by users, such as comments, shares, and reactions (like, wow, etc.).

The experts involved in the annotation have broad experience in corporate communications. Hence, it should be noted that the sets of annotators involved in the two corpora have only a small intersection.

A. Corpus 1: Automobile

The first corpus [7] includes Facebook pages from the automotive, food retail and IT industries. Seven pages from German companies were selected from these industries and their posts were retrieved. The pages include Audi AG, Audi DE, BMW Germany, Edeka, Mercedes Benz, Microsoft DE and Volkswagen DE.

A number of 5,509 posts were examined and annotated by a group of 17 experts. During annotation, the experts evaluated two aspects: the success and a category to which the post can be assigned. The success was either classified as *successful* or *not successful*. The category, on the other hand, describes the topic of the post content. Each post was assigned to exactly one category that best describes its content. The experts defined the following eleven categories for this purpose, based on their experience in corporate communications:

1) Product / Service: Product Launch, Preview, Review
2) Event / Fair
3) Interactions: Contest, survey, question
4) News: News from the environment
5) Entertainment: Memes, Jokes, Virals, Contests
6) Knowledge: Tip, Expertise, Insight, Case Study, FAQ, Research
7) Recruiting / HR: Employee Feature, Interview, Testimonial, Job Advertisement
8) Corporate Social Responsibility (CSR)
9) Advertising / Campaign: testimonial, discounts, lead generation
10) Sponsoring
11) Other category: None of the other categories

B. Corpus 2: Food delivery

The second corpus [8] was created from six of the most important Facebook brand pages of the food delivery industry in Germany. The pages included are Call a Pizza, Deliveroo, Domino's, Lieferando, Mundfein and Smiley's.

About 6,000 of the posts of these pages were annotated by a group of five experts who rated the success of a post as *successful* or *not successful* and assigned one or more out of the eleven categories introduced in Section V-A that best describe its content.

C. Statistics

The two corpora contain about 12,000 annotated posts in total, distributed equally among both corpora. Table II represents the proportion of the posts assigned to each of the categories. It should be noted that each post in the first corpus was assigned to exactly one category, whereas for the second corpus a selection of multiple topics (multi-label) was possible.

TABLE II. Posts per category

	First corpus		Second corpus	
Category	Posts	Percent	Posts	Percent
1. Product/Service	1,300	23.60 %	316	5.22 %
2. Event/Fair	820	14.88 %	368	6.07 %
3. Interactions	1,154	20.95 %	2,370	39.12 %
4. News	441	8.01 %	547	9.03 %
5. Entertainment	372	6.75 %	978	16.14 %
6. Knowledge	355	6.44 %	390	6.44 %
7. Recruiting/HR	75	1.36 %	65	1.07 %
8. Corporate Social Responsibility (CSR)	57	1.03 %	40	0.66 %
9. Advertising/Campaign	605	10.98 %	4,098	67.63 %
10. Sponsoring	88	1.60 %	322	5.31 %
11. Other	242	4.39 %	541	8.93 %

Note: First corpus is single-label while second corpus is multi-label

TABLE III. Posts by success

	First corpus		Second corpus	
	Posts	Percent	Posts	Percent
Not successful	2,126	38.59 %	4,578	76.3 %
Successful	3,383	61.41 %	1,422	23.7 %

In the first corpus, most common classes are *1. Product/-Service*, *3. Interactions* and *2. Event/Fair*, which sum up to almost 60 %. In contrast, the categories *7. Recruiting/HR*, *8. Corporate Social Responsibility* and *10. Sponsoring* appeared only very rarely in the corpus. The distribution of categories in the second corpus differs from that of the first corpus. The three most frequent categories are *9. Advertising/Campaign*, *3. Interactions* and *5. Entertainment*. The less frequent categories were pretty similar to the first corpus, with *7. Recruiting/HR* and *8. Corporate Social Responsibility*.

Table III shows the distribution of success, which was rated with approximately 60 % as successful for the first corpus. This is in contrast to the second corpus where about three-quarters were rated as unsuccessful. Possible reasons for the discrepancy include different conditions among the two corpora, such as the participating annotators, the companies included in terms of industry and size as well as the distribution of the topics among the included posts.

As presented in Table IV, the different topics have varying distribution of success. In addition, each company included in the two corpora also has its individual distribution of success, as shown in Table V. Despite this, we argue that this variety does not have a negative impact on the functionality of the models built with the data, as we show in Section VII.

TABLE IV. SUCCESSFUL POSTS PER CATEGORY

	First corpus	Second corpus
Category	Percent	Percent
1. Product/Service	69.69 %	12.34 %
2. Event/Fair	52.80 %	19.29 %
3. Interactions	67.24 %	35.11 %
4. News	47.39 %	22.12 %
5. Entertainment	62.90 %	48.98 %
6. Knowledge	48.45 %	29.74 %
7. Recruiting/HR	38.67 %	13.85 %
8. Corporate Social Responsibility (CSR)	33.33 %	40.00 %
9. Advertising/Campaign	73.39 %	24.62 %
10. Sponsoring	38.64 %	34.78 %
11. Other	52.48 %	6.84 %

TABLE V. SUCCESSFUL POSTS PER PAGE

First corpus		Second corpus	
Page	Percent	Page	Percent
Audi AG	44.59 %	Call a Pizza	13.86 %
Audi DE	82.19 %	Deliveroo	41.96 %
BMW Germany	69.00 %	Domino's	20.54 %
Edeka	61.09 %	Lieferando	56.63 %
Mercedes Benz	67.22 %	Mundfein	11.71 %
Microsoft DE	14.56 %	Smiley's	8.05 %
Volkswagen DE	67.93 %		

For the annotation of data, the rules of annotation play an important role. Especially when several annotators work together, it has to be ensured that all annotators apply the same rules. A common method for calculating the consistency among annotators is to let all annotators evaluate the same documents and then compare the results. To evaluate the annotations of the first corpus, about ten percent of the posts were annotated a second time by a another expert to identify deviating annotations among experts. An analysis of the double-annotated posts revealed that the two annotators selected success identically in 72 %, while they selected the identical category in only 48 %. We argue that success is much easier to annotate than category, since in certain cases two or even more categories might be applicable to a post. To ensure more consistent annotations, a different annotation method was used for the second corpus, in which the agreement among annotators was calculated using inter-rater reliability. A well-known measure of inter-rater reliability, that can be calculated among two or more annotators, is Fleiss' Kappa [9], that was used for this purpose. The Fleiss' Kappa values are 0.4835 for the topics and 0.6674 for success, which indicates a moderate and substantial agreement [10].

VI. CLASSIFICATION

Following the *AMPEL model creation* shown in Fig. 1 and based on the data sets, the two models for classification by success were trained. The models differ in the point in time at which they are used. The first model allows to predict the success of a post even before its publication, i.e. *a priori*, whereas the second model rates the actual success of a published post retrospectively, i.e. *a posteriori*. The model creation process is based on four steps, we illustrate in Fig. 2.

Fig. 2. AMPEL model training

First, the texts of the posts are subjected to pre-processing. Afterwards, feature extraction is used to derive further features from the data. In a third step, additional features are extracted from the post texts using natural language processing (NLP). Finally, the model creation is carried out.

A. Features

As explained in Section IV, we use different features for the a priori and a posteriori models. Our criterion for the selection of features is the time at which they are available. Under this rule, the feature set for the a priori model only includes the features of a new post at the time it is created. The a posteriori model, on the other hand, contains not only these features but also others that are created after publication, i.e. the interactions.

Table VI provides an overview of the features that we used to build the models. The features can be grouped into three categories:

- The *original attributes*, such as the date of publication, the post type and the text, that come from the data retrieved on the posts. With relation to Table I, the original attributes include the sources *original post* and *interactions*.
- *Annotated attributes*, that were added by the experts during annotation. This is a single category in the case of the first corpus, and one or more in the case of the second corpus.
- *Derived attributes*, that are derived from other original attributes. These include simple ones, for example, the day of week, hour and minute, derived from the date of publication. In addition, there are more complex statistical attributes, that have to be calculated. The POST_LIKES_MAX_RATIO is one of these, which describes the ratio of the likes of the current post to the

likes of the post with the most likes, with only the posts of the corresponding page being included in the calculation.

To make the features usable for the classification models, they had to be encoded. To study the impact of feature selection and encoding on the quality of the results, we created several feature sets represented by columns of Table VI (One-hot, ..., doc2vec) that differ in the composition and encoding of their features. The features of the first model, for predicting success (a priori), are marked with 1, while the features of the second model, for the rating of success (a posteriori), are indicated by 2.

Categorical attributes such as the post type, page, and industry were numerically encoded using *binary encoding* and *one-hot encoding*. The temporal attributes (minute, hour, day, month) were further encoded with sine and cosine. This allows the relationships between times to be correctly represented, such as the fact that the duration between 10 p.m. and 1 a.m. is less than the duration between 3 p.m. and 8 p.m. The text represents an essential content of a post and thus has a significant influence on whether the post will be successful or not. For encoding the texts, the simple techniques bag-of-words (BOW) and term frequency-inverse document frequency (TF-IDF) were used, supplemented by doc2vec [11], as a more advanced technique for word-embeddings. For this purpose, the post texts from the training set served as the basis for the word-embedding. In order to learn more about the content, the texts of the collected posts were analyzed. The objective was to find out which content influences the success of a post so that this information could be transferred into the models used to predict and rate success. For this purpose, the texts were examined in a text analysis for the occurrence of certain influencing factors. The possible influencing factors include hashtags, URLs, emoticons and UTF-8 emojis as an expression of certain emotions, questions that prompt the user to answer, as well as the readability of the text measured by the Flesch formula [12]. Finally, we performed a feature extraction to make the influence factors available as features.

B. Algorithms & Model training

Both models were implemented with several algorithms to optimize the quality of the classification. The selected algorithms include Random Forest and XGBoost [13] from the category of decision trees as well as multilayer perceptron (MLP) as a representative of neural networks. Random Forest is the oldest approach providing a bagging model to compute predictions in parallel over multiple decision trees. The algorithm XGBoost (eXtreme Gradient Boosting) represents a form of gradient tree boosting, which supports sequential linking of decision trees. Both approaches are suitable for a variety of different problems, so we assume that our use case can also benefit from these methods. In addition to these two conventional classification algorithms, models based on neural networks were also tested. For this purpose, we used a simple neural network of type multilayer perceptron (MLP). The network uses a hidden layer of 23 neurons with ReLU activation function and an output layer with the Sigmoid activation function. The MLP was trained over 200 epochs with a batch-size of 50 in a stratified 5-fold cross-validation.

VII. Evaluation

For the evaluation of the algorithms, the data of both corpora was split up by two-thirds into a training set and one-third into a test set. The training set was used to train each of the classifiers. Finally, the classifiers were evaluated using the test set and the metrics Precision, Recall and F_1.

Table VII summarizes the results for the different algorithms. The left half of the table shows the results for the a priori model, while the right half contains the results for the a posteriori model. Both areas are further divided into two sub-areas, for the first and second corpus.

The Precision depicted in Eq. (1) describes the ratio of the true positives, i.e. the actually successful posts, to the true positives and false positives, i.e. all posts classified as successful. Since we aim towards correct predictions and ratings, this metric represents the most important one for our use case. For both corpora, the Precision for the a posteriori model turns out to be better than that for the a priori model. This was to be expected since the success of a post correlates with high values for the interactions. In the a priori model, on the other hand, a more causal relationship between other features and success must be identified.

$$\text{Precision} = \frac{tp}{tp + fp} \tag{1}$$

The Recall depicted in Eq. (2) describes the ratio of the true positives, i.e. the posts that are correctly classified as successful, to the sum of all actually successful posts, i.e. the true positives and false negatives. The Recall is less important than Precision for our use case as we aim to minimize the rate of false positives. Nevertheless, it could be of interest for certain use cases, e.g. the identification of new topics that might lead to a successful post. As with Precision, the Recall for both corpora tends to be better for the a posteriori model than for the a priori model for the same reasons. The Recall of the a priori model for the second corpus is rather low, while that of the other three models is good.

$$\text{Recall} = \frac{tp}{tp + fn} \tag{2}$$

Overall, we consider the values for both corpora to be good, whereby the Precision for the models of the second corpus are higher than those for the first corpus. For Recall, on the other hand, the situation is the opposite, with better values for the first corpus. Looking at F_1 depicted in Eq. (3), a fairly consistent picture emerges across the models.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{tp}{tp + \frac{1}{2}(fp + fn)} \tag{3}$$

To better understand the results, we calculated the feature importance for the models based on the XGBoost algorithm. The values we used represent the gain of the features. In the a

TABLE VI. FEATURES

Attribute			Feature set					
Name	Type	Derived from	One-hot	Binary	Cyclic times	BOW	TF-IDF	doc2vec
Category	A		1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Date of publication	O							
Post type	O							
Post type (binary)	D	Post type		1, 2				
Post type (one-hot)	D	Post type	1, 2		1, 2	1, 2	1, 2	1, 2
Page of post	O							
Page of post (binary)	D	Page of post		1, 2				
Page of post (one-hot)	D	Page of post	1, 2		1, 2	1, 2	1, 2	1, 2
Industry of page	O							
Industry of page (binary)	D	Industry of page		1, 2				
Industry of page (one-hot)	D	Industry of page	1, 2		1, 2	1, 2	1, 2	1, 2
Year	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Month	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Day	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Hour	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Minute	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Day of week	D	Date of pulication	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Minute sinus	D	Minute			1, 2			
Minute cosinus	D	Minute			1, 2			
Hour sinus	D	Hour			1, 2			
Hour cosinus	D	Hour			1, 2			
Day sinus	D	Day			1, 2			
Day cosinus	D	Day			1, 2			
Month sinus	D	Month			1, 2			
Month cosinus	D	Month			1, 2			
Text	O							
Text (BOW)	D	Text				1, 2		
Text (TF-IDF)	D	Text					1, 2	
Text (doc2vec)	D	Text						1, 2
Text length	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text complexity (Flesch)	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains URLs	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains questions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Number of hashtags in text	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains affection emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains happy emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains skeptical emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains sad emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains angry emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains surprise emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Text contains unwell emotions	D	Text	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
Number of likes	O		2	2	2	2	2	2
Number of haha	O		2	2	2	2	2	2
Number of wow	O		2	2	2	2	2	2
Number of angry	O		2	2	2	2	2	2
Number of love	O		2	2	2	2	2	2
Number of sad	O		2	2	2	2	2	2
Number of comments	O		2	2	2	2	2	2
Number of shares	O		2	2	2	2	2	2
POST_LIKES_MAX_RATIO	D	Number of likes	2	2	2	2	2	2
SHARES_MAX_RATIO	D	Number of shares	2	2	2	2	2	2
COMMENTS_MAX_RATIO	D	Number of comments	2	2	2	2	2	2
ANGRY_MAX_RATIO	D	Number of angry	2	2	2	2	2	2
LOVE_MAX_RATIO	D	Number of love	2	2	2	2	2	2
HAHA_MAX_RATIO	D	Number of haha	2	2	2	2	2	2
WOW_MAX_RATIO	D	Number of wow	2	2	2	2	2	2
SAD_MAX_RATIO	D	Number of sad	2	2	2	2	2	2
POST_LIKES_MEDIAN_RATIO	D	Number of likes	2	2	2	2	2	2
SHARES_MEDIAN_RATIO	D	Number of shares	2	2	2	2	2	2
COMMENTS_MEDIAN_RATIO	D	Number of comments	2	2	2	2	2	2

1. Feature of *a priori* model, 2. Feature of *a posteriori* model, A. Annotated attribute, D. Derived attribute, O. Original attribute

posteriori model for both corpora, the assumed importance of the interactions was confirmed, as the various interactions had by far the highest importance, followed by the features derived from interactions, such as POST_LIKES_MEDIAN_RATIO. Also important features are the company behind the page and its industry, as well as the category and type of the post. For the a priori model, on the other hand, the picture is more diverse. Here, the company behind the page and their industry are the most important features. Also important are the category and post type, certain keywords from the text, as well as the time, encoded with sine and cosine. These are followed by other highly weighted features derived from the text, such as the occurrence of questions, URLs, hashtags and emoticons, but also the simple text length and text complexity according to Flesch.

VIII. Prototype

To evaluate the technical feasibility of our idea, we developed a prototype that implements the workflow from Section III. The prototype is realized as a web service that can be used by social media managers to draft posts and predict their future success. In the background, the web service communicates via a REST interface with the prediction service, providing the classification of post success. After success prediction, the user can publish the post directly on Facebook from the tool. In Fig. 3 the user interface of the web service for post creation is shown. The posting text, its category and the time of publication can be adjusted and evaluated via the user interface until the author is confident with the predicted result calculated by the classifier. The result is transformed into a three-value representation in shape of a traffic light (in German: Ampel). A green light means that the post was predicted to be successful ($success_{pred} > \frac{2}{3}$), while a red light indicates a post might be not successful ($success_{pred} \leq \frac{1}{3}$). If the success classification is in the range between successful and not successful ($\frac{1}{3} < success_{pred} \leq \frac{2}{3}$), a yellow light is shown. Following the writing and evaluation, the drafted post can be saved for automatic publication at the desired time.

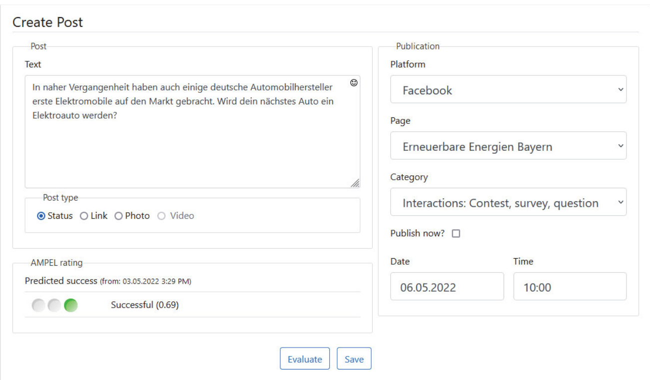

Fig. 3. Prototype with prediction of success

The example shown in Fig. 3 contains the German translation of the text *"In the near past, also several German car manufacturers have launched their first electric cars on the market. Will your next car be an electric car?"* with the aim to publish it on a page for renewable energies on 06/05/2022 at 10 AM. According to the AMPEL prototype, the user is predicted that his/her post will be successful, with a probability of almost 70 %.

In addition to the functionality to rate and publish new posts, the prototype also offers the functionality to analyze posts already published. With the help of various curves and diagrams, the actual success of postings can be monitored. This also allows an author to compare the predicted success ratings with the actual success achieved. For this purpose, the prototype retrieves the published posts every 15 minutes. On every change of a posts interactions, its actual success will be re-evaluated based on the a posteriori model.

IX. Related Work

Cvijikj et al. [14] study how certain characteristics of posts on a brand page affect the number and timing of user interactions. The characteristics examined, included the categories, post types and publishing times of 120 posts from the Facebook page of a Swiss consumer goods brand. In order to evaluate the effect of the category, each post was assigned to one of seven categories. The results show a significant impact of post type and category on the number of likes and comments. These findings were confirmed as part of a broader study [3], where the authors examined 14 brand pages on Facebook with a similar methodology.

A trend detection system [2] employs TF-IDF to automatically categorize posts into one of three categories of trending topics. To do this, it identifies the most relevant terms in the post text and assigns them to a topic category based on their distribution and co-occurrence.

While the studies above examined which elements of a post influence user interactions, Moro et al. [1] attempt to predict the impact of posts on the number of user interactions. The data used included 790 Facebook posts and their attributes (type of post, number of views, etc.) from a cosmetics company. Based on this data, a support vector machine was used to predict various metrics.

To predict the popularity of Instagram posts with different content (images, videos, etc.), a regression based on user and post features can be performed [15]. In addition, statistical features were used in this study to predict the log-normalized number of likes. The work of Carta et al. [16] made popularity predictions on Instagram to determine whether a post will have a positive or negative deviation from the average number of likes, using a binary classification task, as in our use case. They used XGBoost and Random Forest for their classifications.

Abousaleh et al. [17] attempted to predict the popularity of images on Flickr by combining visual features from the images with social features inferred from the user, the post's metadata, and the time of publication. They trained two separate convolutional neural networks (CNNs) based on the visual and social features and finally combined their results.

TABLE VII. RESULTS OF THE CLASSIFICATION

		A priori						A posteriori					
		First corpus			Second corpus			First corpus			Second corpus		
Meas.	Attr.	XGB	RF	MLP	XGB	RF	MLP	XGB	RF	MLP	XGB	RF	MLP
Prec.	One-hot	0.7350	0.6933	0.7401	0.7287	0.7729	0.6195	0.8270	0.8253	0.8334	0.9185	0.9298	0.8444
	Binary	0.7398	0.7142	0.7080	0.7295	0.7557	0.5527	0.8311	0.8312	0.8076	0.9049	0.9210	0.8860
	Cyclic times	0.7336	0.6952	**0.7601**	0.7135	0.7706	0.6276	0.8364	0.8239	**0.8438**	0.9165	0.9275	0.8471
	BOW	0.7420	0.6194	0.7255	0.7131	**0.8667**	0.4364	0.8368	0.6428	0.7530	0.9187	0.9877	0.8454
	TF-IDF	0.7381	0.6214	0.7307	0.6992	0.8462	0.5159	0.8333	0.6425	0.8160	0.9233	**0.9886**	0.8868
	doc2vec	0.7363	0.6908	0.7099	0.7227	0.7826	0.6313	0.8343	0.8265	0.8412	0.9204	0.9271	0.9080
Rec.	One-hot	0.8606	0.9455	0.8406	**0.5661**	0.4360	0.5375	0.8680	0.8771	0.8866	0.7686	0.7665	0.8276
	Binary	0.8515	0.9134	0.9085	0.5517	0.4091	0.3397	0.8729	0.8738	0.8967	0.7665	0.7707	0.8060
	Cyclic times	0.8564	0.9464	0.7868	**0.5661**	0.4442	0.5024	0.8688	0.8721	0.8722	0.7707	0.7665	**0.8304**
	BOW	0.8828	**0.9975**	0.7751	0.5496	0.0269	0.4803	0.8754	**0.9934**	0.8969	0.7707	0.1653	0.7051
	TF-IDF	0.8672	0.9967	0.8193	0.5331	0.0227	0.4482	0.8705	**0.9934**	0.8447	0.7707	0.1798	0.7208
	doc2vec	0.8548	0.9530	0.9079	0.5599	0.4091	0.4419	0.8680	0.8762	0.8700	0.7645	0.7624	0.7693
F_1	One-hot	0.7929	0.8000	0.7830	**0.6372**	0.5575	0.5557	0.8470	0.8504	**0.8560**	0.8369	**0.8403**	0.8245
	Binary	0.7917	0.8016	0.7917	0.6282	0.5308	0.4072	0.8515	0.8520	0.8468	0.8300	0.8391	0.8344
	Cyclic times	0.7903	0.8015	0.7688	0.6313	0.5636	0.5368	0.8523	0.8473	0.8544	0.8373	0.8394	0.8278
	BOW	**0.8063**	0.7642	0.7448	0.6208	0.0521	0.4418	0.8556	0.7806	0.8152	0.8382	0.2832	0.7549
	TF-IDF	0.7974	0.7655	0.7682	0.6049	0.0443	0.4621	0.8515	0.7803	0.8276	0.8401	0.3042	0.7827
	doc2vec	0.7911	0.8010	0.7926	0.6310	0.5373	0.4966	0.8508	0.8506	0.8522	0.8352	0.8367	0.8207

XGB: XGBoost, RF: Random forest, MLP: Multilayer perceptron

X. DISCUSSION

The AMPEL models presented in this paper are specialized for the industries from the two corpora used. This is in contrast to many other models where the focus is on generalization. Our approach is intended to support companies of selected industries in the creation of their posts. Hence, we argue that in order to achieve high quality, AMPEL models should always be trained and evaluated specifically for the industries they are used for. Moreover, the approach can be adapted to the post behavior of other languages, cultures or social media portals.

Since the post text is a significant criterion for the success of a post, it may be advisable to analyze it even more deeply. For this purpose, BERT [18] offers a variety of pre-trained models specialized in processing texts of different languages. This, however, does not make the evaluation of non-textual features obsolete. the emotional impact of images on people is great, which is why they are also frequently used in social media. This makes it interesting to include the content of images as a classification feature. For this purpose, there exist some frameworks from the fields of object recognition and image classification, which are mostly trained using neural networks. The presented AMPEL approach does not include such features yet, but provides the infrastructure to integrate them.

The user interface depicted in Fig. 3 shows that the users of the AMPEL prototype have to select the category of the drafted text themselves. This way of entering information is prone to errors, which can result in an incorrect classification of the posts. This problem can be counteracted by automating the detection of the post category based on the post text. For this purpose, the use of a zero-shot classifier would be conceivable.

In order to be able to monitor the actual success of the posts in a timely manner, changes on posts have to be retrieved with as little delay as possible. Since data from social media cannot be retrieved as often as desired, due to rate-limits of the platforms, the current solution of checking every post for updates every 15 minutes is not optimal. With an increasing number of pages and posts, intelligent polling algorithms are required to keep the data up to date with as few well-timed requests as possible [19].

As previously discussed, the patterns of communication on social media evolve over time, making it necessary to stay up to date with AMPEL models. This can be a costly process as it requires continuous labeling by annotators. To make long-term service more efficient, AMPEL models could be turned into a self-learning system. A good starting point for this could be to use co-training [20], which allows the training data to be gradually expanded by data that have not been labeled yet. In addition, transfer learning approaches such as few-shot classification could be used to keep the system up to date.

XI. CONCLUSION

In this paper, we presented the AMPEL approach, that allows to predict and rate the success of posts for company pages on Facebook. The prediction can be run while creating posts and assist in selecting its content and publication date. In addition, a prototype for creating and evaluating posts was presented. The implementation enables not only experts, but also inexperienced users to create successful posts. We have also revealed that success depends on certain features. Using this knowledge, we trained two models based on different features and algorithms that allow to predict and rate the success of posts. The results of the evaluation showed that both models provide good results. The best overall performance was achieved by classifiers based on XGBoost.

As a next step, a user study is needed to measure and evaluate how the presented approach can support users improving the success of their posts for brand pages. For this purpose, the two introduced prediction models can be extended to include posts from particular industries. In addition, the AMPEL approach allows the processing of postings from other social media platforms, such as YouTube or Twitter, to be integrated.

XII. Acknowledgment

The presented work was conducted as part of a project funded by *Forschungs- und Entwicklungsprogramm Informations- und Kommunikationstechnik des Freistaates Bayern*. Funding reference number: IUK482/002. We would like to thank our students Daniel Stiller and Karim Belkhiria, who have contributed to this work with their final theses.

References

[1] S. Moro, P. Rita, and B. Vala, "Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach," *Journal of Business Research*, vol. 69, no. 9, pp. 3341–3351, 2016.

[2] I. P. Cvijikj and F. Michahelles, "Understanding social media marketing," in *Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments*, A. Lugmayr, H. Franssila, C. Safran, and I. Hammouda, Eds. New York, NY: ACM, 2011, p. 175.

[3] I. Pletikosa Cvijikj, E. D. Spiegler, and F. Michahelles, "The effect of post type, category and posting day on user interaction level on facebook," in *IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT), 2011 and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. Piscataway, NJ: IEEE, 2011, pp. 810–813.

[4] B. Wu, T. Mei, W.-H. Cheng, and Y. Zhang, "Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[5] M. S. Fiebert, A. Aliee, H. Yassami, and M. D. Dorethy, "The life cycle of a facebook post," *The Open Psychology Journal*, vol. 7, no. 1, pp. 18–19, 2014.

[6] F. Erlandsson, R. Nia, M. Boldt, H. Johnson, and S. F. Wu, "Crawling online social networks," in *2015 Second European Network Intelligence Conference (ENIC 2015)*. Piscataway, NJ: IEEE, 2015, pp. 9–16.

[7] M.-E. Keller, B. Stoffelen, D. Kailer, P. Mandl, and J. Althaller, "Predicting the success of posts for brand pages on facebook," in *Proceedings of the 17th International Conference WWW/Internet 2018, Budapest, Hungary: IADIS*, 2018.

[8] M.-E. Keller, J. Forster, P. Mandl, F. Aich, and J. Althaller, "A german corpus on topic classification and success of social media posts," in *Proceedings of the 25th Conference of Open Innovations Association FRUCT*, ser. FRUCT'25. Helsinki, Finland: FRUCT Oy, 2019.

[9] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *0033-2909*, vol. 76, no. 5, pp. 378–382, 1971.

[10] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, p. 159, 1977.

[11] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the 31st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, Eric P. Xing and Tony Jebara, Eds., vol. 32. Bejing, China: PMLR, 2014, pp. 1188–1196. [Online]. Available: http://proceedings.mlr.press/v32/le14.html

[12] R. Flesch, "A new readability yardstick," *The Journal of applied psychology*, vol. 32, no. 3, pp. 221–233, 1948.

[13] T. Chen and C. Guestrin, "Xgboost," in *KDD2016*, B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, and R. Rastogi, Eds. New York, NY: Association for Computing Machinery Inc. (ACM), 2016, pp. 785–794.

[14] I. Pletikosa Cvijikj and F. Michahelles, "Monitoring trends on facebook," in *IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC), 2011*, J. Chen, Ed. Piscataway, NJ: IEEE, 2011, pp. 895–902.

[15] M. Gayberi and S. G. Oguducu, "Popularity prediction of posts in social networks based on user, post and image features," in *Proceedings of the 11th International Conference on Management of Digital EcoSystems*, R. Chbeir, Y. Manolopoulos, S. Ilarri, and A. Papadopoulos, Eds. New York, NY, USA: ACM, 2019, pp. 9–15.

[16] S. Carta, A. S. Podda, D. R. Recupero, R. Saia, and G. Usai, "Popularity prediction of instagram posts," *Information*, vol. 11, no. 9, p. 453, 2020.

[17] F. S. Abousaleh, W.-H. Cheng, N.-H. Yu, and Y. Tsao, "Multimodal deep learning framework for image popularity prediction on social media," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 3, pp. 679–692, 2021.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North*, J. Burstein, C. Doran, and T. Solorio, Eds. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186.

[19] M.-E. Keller, A. Döschl, P. Mandl, and A. Schill, "When she posts next? a comparison of refresh strategies for online social networks," in *The 23rd International Conference on Information Integration and Web Intelligence*, E. Pardede, M. Indrawan-Santiago, P. D. Haghighi, M. Steinbauer, I. Khalil, and G. Kotsis, Eds. New York, NY, USA: ACM, 2021, pp. 123–129.

[20] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory - COLT' 98*, P. Bartlett and Y. Mansour, Eds. New York, New York, USA: ACM Press, 1998, pp. 92–100.