Identifying FinTech Innovations Using BERT

Doina Caragea
Department of Computer Science
Kansas State University
Manhattan, KS
dcaragea@ksu.edu

Mark Chen
J. Mack Robinson College of Business
Georgia State University
Atlanta, GA
machen@gsu.edu

Theodor Cojoianu
Michael Smurfit Graduate Business School
University College Dublin
Dublin, Ireland
theodor.cojoianu@ucd.ie

Mihai Dobri
Department of Computer Science
Kansas State University
Manhattan, KS
mihaidobri@ksu.edu

Kyle Glandt
Department of Computer Science
Kansas State University
Manhattan, KS
kglandt@ksu.edu

George Mihaila
Department of Computer Science and Engineering
University of North Texas
Denton, TX
GeorgeMihaila@my.unt.edu

***Abstract*—Advancements in technology have resulted in the emergence of numerous FinTech innovations. However, a global understanding of such innovations is limited, due to a lack of an underlying taxonomy and benchmark datasets in the FinTech domain. To address this limitation, we develop a FinTech taxonomy and manually annotate a set of FinTech patent abstracts according to the taxonomy. We use the annotated dataset to train deep learning models, specifically recurrent neural networks and convolutional neural networks combined with state-of-the-art BERT transformers. Experimental results show that the deep learning models can accurately identify FinTech innovations. We use our best performing BERT-based model on a large dataset of financial patent abstracts, and shortlist a set of 25,580 FinTech patent applications submitted to the European and US Patent Offices between 2000 and 2017. We illustrate how an analysis of the shortlisted set can be used to gain understanding of what FinTech innovations are, where and when they emerge, and provide the basis for further work on what their impact is on the companies investing in them, and ultimately on society.**

***Index Terms*—FinTech, Financial Technologies, Patent Classification, Deep Learning, BERT**

I. Introduction

The financial and technology sectors have been interwoven in the last 150 years [1], ever since the communication infrastructure underpinning financial transactions has been built. Post-2008, the year marking the all-time low trust in financial institutions, advancements in technology and data science have resulted in the emergence of numerous FinTech start-ups and placed start-ups, as well as highly trusted technology sector incumbents in a good position to challenge the traditional financial sector in the provision of financial services [2, 3]. According to KPMG [1], in 2019, global investments in FinTech start-ups attracted $135.7B with 2,693 deals, most notably in FinTech sub-sectors such as payment technologies and investment and lending platforms. FinTech innovations do not occur only in start-ups. On the contrary, financial sector incumbents are responding in numerous ways to the technological disruption overtaking the sector. Moreover, technology companies, which have been historically close to consumers, are also emerging as providers of financial services. In the US, in 2016 alone, JP Morgan has spent more than $9.5 billion in revamping its IT infrastructure, out of which $600 million was spent on developing FinTech solutions, either in-house or through partnerships[2].

Despite significant investments in FinTech solutions, the literature to date has been limited in explaining what FinTech innovations are, where and why they emerge, and what is their impact on society and on the financial performance of companies that invest in them [4], in part due to a lack of a widely-accepted FinTech innovations taxonomy and datasets categorized according to such taxonomy. To fill in this gap, we aim to study the global landscape of FinTech innovations starting from a global patent dataset of over 100 million patent applications published between 2000 and 2017. Towards this goal, we first create a FinTech innovation taxonomy corroborated from an extensive literature search on working papers and published academic articles, reports and materials related to FinTech. We use a list of financial terms to pre-filter financial-related patents. We then build a substantial corpus of manually labelled FinTech patents, and use it to train and evaluate different types of deep learning classifiers, with focus on BERT models [5]. Finally, we run our best performing classifier on the abstracts of patent applications pre-filtered using financial terms and arrive at our global dataset of FinTech innovations. Using the resulting dataset, we provide a spatial and temporal distribution of FinTech innovation emergence across different FinTech categories.

The rest of the paper is organized as follows: We discuss related work in Section II. The FinTech dataset is described in Section III, while the models we studied are introduced in Section IV. We describe our experimental setup in Section V. We discuss the results of the experiments in Section VI, and perform a large-scale spatial and temporal analysis of FinTech

[1]https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/02/pulse-of-fintech-h2-2019.pdf

[2]https://www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/investor-relations/documents/2016-annualreport.pdf

innovations in Section VII. Finally, we conclude the paper and present ideas for future work in Section VIII.

II. Related work

Over the past decade, academic research in FinTech has grown in tandem with the exponential rise of new FinTech start-ups around the world [6]. Several journals have hosted special topics dedicated to FinTech Innovations, including the Review of Financial Studies [7] and the Journal of Management Information Systems [8]. However, very few studies have been able to provide a systematic overview of FinTech innovations, partly due to a lack of an internationally recognised taxonomy, and partly due to a lack of datasets that would enable large-scale analysis using machine learning and deep learning approaches. Such taxonomies and datasets are available for general innovations and have been used successfully to automatically classify patents and gain insights into general innovations trends in the last decade [9, 10, 11, 12, 13, 14, 15, 16, 17].

In the FinTech area, one of the first studies to use machine learning to identify FinTech innovations, and the implications on the financial performance of companies who invest in such innovations, was performed by Chen et al. [4]. The authors employed text-based machine learning approaches to classify and analyze innovations according to their key underlying technologies. Chen et al. [4] used a dataset of US patents covering years 2003-2017, pre-filtered using 487 financial terms. A subset of 1,800 patents was manually annotated according to 9 categories. These categories include 7 FinTech categories (specifically, *Cyber-security*, *Mobile Transactions*, *Data Analytics*, *Blockchain*, *Peer-to-peer*, *Robo-adviser* and *Internet of Things*), a category for financial patents that are not FinTech, and a category for non-financial patents. The manually annotated dataset was used to train and evaluate several machine learning classifiers. Empirical results showed that an ensemble classifier, consisting of linear support vector machines (SVM), Gaussian SVM, and neural network models trained on patent text, performed the best, with an accuracy of 82.6% and an F1 score of 76.3%.

In another recent study, Xu et al. [18] trained random forest (RF) classifiers (which can be seen as ensembles of decision trees) to identify FinTech patents. The original dataset used in their study was extracted from the Lens database, and covered years 2014-2018. A set of 478 financial terms was used to filter financial innovations. A subset of 1,800 patents was manually annotated according to 9 categories, including 7 FinTech categories (*Encryption & Security*, *Mobile Payments*, *Big Data Analytics*, *Blockchain*, *Online Lending*, *Expert Advisor*, and *Internet of Things*), and the 2 additional categories from [4]. The labeled subset was used to train and evaluate RF classifiers. Empirical results showed that the best performing classifier achieved an average accuracy of 71.67%.

While the datasets used by Chen et al. [4] and Xu et al. [18] do not include exactly the same categories, they are based on similar raw data (i.e., patent applications filed by inventors, in patent offices such as USPTO or EPO). The characteristics of the two datasets are summarized in Table I, and some of them are discussed below:

- Both datasets consist of filtered patents with legal jurisdiction in the US, and belong to the G&H classes from the International Patent Classification (IPC) hierarchy.
- Both [4] and [18] used similar "lists of financial terms" consisting of 487 and 478 terms, respectively, to filter patents potentially related to financial services.
- Both studies identified 7 FinTech categories, and 2 additional categories to capture not FinTech, and non-financial patents, respectively. Chen et al. [4] identified the seven FinTech categories based on insights from a general reading of FinTech reports and articles. Xu et al. [18] selected their seven FinTech categories based on a Financial Stability Board (FSB) report from 2017.
- Both studies manually labeled small subsets of patents (specifically, 1,800 patents) according to the 9 categories considered. Chen et al. [4] labeled 200 patents in each of the 9 categories considered, while Xu et al. [18] selected a random sample of 1800 patents and labeled them according to the 9 categories.

The prior works on FinTech innovation classification [4, 18] have employed traditional machine learning approaches, and have found that ensemble-type approaches show promising results. However, in the light of the growing success that deep learning approaches have seen in recent years, several works [13, 14, 15, 16, 17] have used such approaches to automatically classify general patents according to standard categories in the International Patent Classification (IPC) or the Cooperative Patent Classification (CPC) taxonomies, and to improve the overall financial technology solutions.

For example, [12, 14] used recurrent neural networks, specifically, long short-term neural networks (LSTM) [19], together with word2vec embeddings [20], to classify patents into IPC categories, while [21] used gated recurrent unit (GRU) networks, together with fastText embeddings [22] for the same task. Similarly, [13, 16] used word embeddings, including Word2vec [20] and GloVe [23], together with convolutional neural networks (CNN) for text classification [24]. Hu et al. [15] build a hierarchical feature model that combined CNN and bidirectional LSTM (bi-LSTM) networks to capture both local lexical-level features and global sequential dependencies. The authors showed that the combined model achieved better performance than the independent CNN and LSTM/Bi-LSTM models on mechanical patent documents. Finally, Lee and Hsiang [17] obtained state-of-the-art results with BERT models [5] on the task of classifying patent documents according to the IPC or CPC taxonomies. Specifically, [17] used large datasets of patent documents to fine-tune a pre-trained BERT-base model on the general patent classification task.

Beyond simply improving performance on the task of automatic patent classification, it is also of interest to analyze innovation trends, as cutting-edge technologies are permanently pioneered by scientists across the world. For example, trends within the technology domain, or within the financial services

	Dataset Characteristics	Chen *et al.* [4]	Xu *et al.* [18]	Our dataset
(1)	Source of patents	BDSS	Lens	**Orbis/Patsat**
(2)	Years covered	2003-2017	2014-2018	**2000-2017**
(3)	Legal jurisdiction of patents	US	US	**US + Europe**
(4)	IPC classes used	G&H	G&H	**G&H**
(5)	Initial number of patents based on criteria (1)-(4) above	200151	1181162	**6.8M**
(6)	Financial terms for filtering financial patents	487	478	**516**
(7)	Number of patents after filtering with financial terms (6)	67948	37156	**38228**
(8)	Number of FinTech categories considered	7	7	**5**
(9)	Number of manually annotated patents used for training	1800	1800	**2350+1500**
(10)	Total number of FinTech patents identified out of (7)	6511	3602	**25580**

TABLE I
CHARACTERISTICS OF TWO EXISTING FINTECH DATASETS USED BY CHEN *et al.* [4] AND XU *et al.* [18] BY CONTRAST WITH OUR DATASET.

sector, can be foreseen by analyzing patent applications and patent grants. Chae and Gim [25] proposed a model based on the existing patent classification schemes, IPC and CPC, to extract accurate innovation trends, as well as common invention patterns. Sofean et al. [26] emphasized the need to predict proper relationships and trends among technological areas of inventions, and proposed to use natural language processing techniques to achieve this task. In the FinTech area, Chen *et al.* [4] used the results of their best ensemble classifier on a large set of financial patents to identify temporal trends with respect to the category of the innovation technology or with respect to its author. Furthermore, they used the results to estimate the value of a category by considering stock price responses, and showed that the *Internet of Things*, *Robo-advising*, and *Blockchain* categories generate the most significant financial gains to companies that invest in them.

To advance the research on identifying FinTech innovation started by Chen et al. [4] and Xu et al. [18], in this work, we first refine the list of financial terms provided by Chen et al. [4], and use them to identify patents, with jurisdiction in both the US and Europe, potentially related to financial services. Furthermore, we propose a new, improved FinTech taxonomy consisting of 5 categories (specifically, *Insurance*, *Payments*, *Investments*, *Fraud*, and *Data Analytics*), and manually label 2,530 according to these categories. In addition, we label 1,500 *Non-FinTech* patents. Given the success of the deep learning approaches on general patent classification tasks, and in particular, the state-of-the-art results produced by BERT models, we focus on classifying FinTech patents using BERT-type models and compare the results with results obtained with LSTM and CNN models, which take BERT embeddings as input. We also compare the results of the deep learning models with the results of traditional machine learning approaches that have been successfully used on FinTech data in prior work, specifically ensemble classifiers [4] and random forests [18]. Finally, we use the results of the best BERT models to classify a large set of financial-related patents, and perform a spatial-temporal analysis based on the classification results to gain insights into the landscape of the FinTech innovation.

III. FINTECH DATASET

In this section, we describe the process we followed to create our FinTech dataset and taxonomy, as well as the annotation of the subset used for training machine learning and deep learning models.

Fig. 1. WordCloud of the 100 most frequent financial terms in the set of 38,228 patent documents used in our analysis. The size of a term is proportional with the number of patents in which the term appears.

A. FinTech Dataset Construction

We retrieve all the patents filed between 2000 and 2017 from the matched Orbis-PATSTAT database, which contains bibliographical entries for over 100 million patent documents filed around the world, as well as details about their corporate ownership. From these, we select the patents developed/owned by corporate entities, and further filter only those with IPC codes pertaining to the fields G and H, which cover innovations related to digital computing, including many FinTech categories [4]. This filtering process results in 6.8 million patents. Following [4], we start with a dictionary of 487 financial terms from the Campbell R. Harvey's Hypertextual Finance Glossary and the online Oxford Dictionary of Finance and Banking, and develop an enhanced set of 516 financial terms. We use these terms to select patents that contain at least one keyword from the set. The number of patents that contain at least one of the

516 financial terms is 38,228. The distribution of the 100 most frequent financial terms in our dataset is illustrated in Fig. 1, where the size of each term is proportional to the number of patent documents in which that term appears. We use this resulting dataset of 38,228 potential FinTech patents in our analysis as outlined below. Characteristics of our dataset are summarized in the last column of Table I, by contrast with the characteristics of the prior datasets [4, 18]. It is worth noting that our initial dataset is the largest among the three, as it covers both US and European patents. However, our dataset of potential FinTech patents is smaller than the one in [4], where a smaller number of filtering terms was used. The reason for this may be that we filter out duplicate patents that are filed in different jurisdictions under different application numbers.

B. FinTech Innovation Taxonomy

There is a wide range of financial products and services that fall under the FinTech umbrella. Currently, there is no comprehensive, well-accepted taxonomy to analyse the sector. Hence, we build a FinTech taxonomy by corroborating taxonomies which emerged from our research of numerous articles, reports and market maps from both academia and industry [4, 27, 28, 29, 30, 31, 32, 33]. Our taxonomy aims to capture innovations that pursue the integration of more sophisticated IT tools and data science solutions in financial products. It contains five FinTech categories, specifically, *Data Analytics*, *Fraud*, *Insurance*, *Investments* and *Payments*. Applications corresponding to these categories, together with an example of a patent filling abstract in each category are shown in Table II. Our FinTech taxonomy is aligned with that of [4]. However, in some respect, our taxonomy is more general as we include a broader range of FinTech innovations, but in other respects, we exclude some applications which are not necessarily specific to the financial sector, included [4] (e.g., *Blockchain* and *Internet-of-Things*).

C. FinTech Dataset Annotation

To be able to train machine learning and deep learning models for FinTech patent identification, we manually annotated/labeled a subset of our patent dataset. Specifically, we manually labeled 500 patents in each of the following categories: *Fraud*, *Insurance*, *Investments* and *Payments*, and 350 patents in the *Data Analytics* category (the number of manually labeled patents in this category is smaller as these patents were more difficult to identify during the manual analysis). Furthermore, we manually labeled a subset of 1,500 Non-FinTech patents. Thus, together our manually labeled dataset contains 2,350 FinTech patents and 1,500 Non-FinTech patents, for a total of 3,850 manually labeled patents. To train and evaluate our models, the dataset was split into training and test subsets, where the training subset contains 80% of the labeled data and the test dataset contains 20% of the data[3].

We used the best performing model to subsequently identify FinTech patents in the set of 38,228 patents that we assembled (except for those that were manually annotated). Finally, we shortlist a set of 25,580 FinTech innovations. We should note that the ratio of FinTech innovations is higher in our dataset as compared to [4, 18], potentially due to different jurisdictions captured in our study. Our final dataset was used to illustrate the spatial and temporal distribution of FinTech innovations filed by companies from all around the world at the European and US Patent offices. Throughout this process, we use the abstract text of each patent as input to the models.

[3]To enable progress in this area, our annotated dataset (in the form of patent identification number and label) will be made publicly available.

IV. BERT-based Models

In this section, we describe the deep learning models that we use in our analysis of FinTech patents.

A. BERT Model

We use BERT, which stands for Bidirectional Encoder Representations from Transformers [5], as the core approach for the task of classifying FinTech patent documents, given that BERT models have produced state-of-the-art results for many text classification tasks [34], including classification of general patent documents [17]. BERT is a language model that uses a deep bidirectional transformer encoder architecture [35] to encode sentences and their tokens into dense vector representations. A generic model is pre-trained on a large corpus of un-annotated text (e.g., Wikipedia) using two self-supervised learning tasks: masked word prediction (a.k.a., masked language modeling, or MLM) and next sentence prediction (NSP). BERT takes as input a sequence of word tokens, where the first token is a special token denoted by [CLS] (the output representation of the [CLS] token can be seen as a semantic representation for the whole input sequence). A BERT input sequence consists of one or two sentences. For an input sequence consisting of two sentences, the two sentences are separated by another special token, denoted by [SEP]. Embeddings of the input tokens are provided to a multi-layer bidirectional transformer encoder, which transforms the original input embeddings into contextual output embeddings. Fig. 2 shows the architecture of a generic BERT model, which takes two sentences as input.

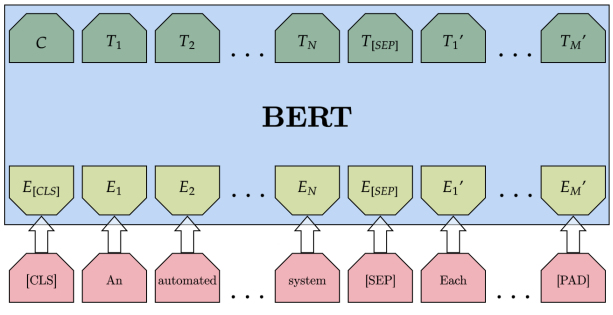

Fig. 2. BERT architecture used for pre-training (figure adapted from [5]).

A generic BERT model, pre-trained on a large corpus, can be further *pre-trained* and/or *fine-tuned* for specific NLP tasks [5]. Particularly, a BERT model for the FinTech patent

Fintech category	Applications	Examples of patent filing abstracts
Data Analytics	**Software, data infrastructure and analytics for financial services**	"[. . .] computer programs product are provided for automated generic and parallel aggregation of characteristics and key figures of unsorted mass data being of specific economic interest, particularly associated with financial institutions, and with financial affairs in banking practice."
Fraud	**Fraud detection, security infrastructure, identity verification & compliance**	"This invention provides a system and method for reducing the fraud related to remittance transactions initiated at web portals. [..] For example, a funding agency computer that enables a remittance transaction can request that a mobile platform computer verify a customer with a mobile personal identifier. The mobile platform computer can request the mobile personal identifier from a customer via the customer's mobile handset device."
Insurance	**Life, general & (re)insurance software analytics**	"An automated assignment system may operate with a computer to automatically assign insurable events to one or more organizational entities associated with an insurance organization. The automated assignment system may categorize the insurable event."
Investments	**Portfolio management, lending and investing platforms and portfolio analytics**	"A visual interactive multi-criteria decision-making method and computer-based apparatus for portfolio management. The method/apparatus supports partitioning of a portfolio of physical or other assets into two mutually exclusive categories, such as assets recommended for sale and assets recommended for retention."
Payments	**Mobile payments & transfer**	"A mobile payment platform and service provides a fast, easy way to make payments by users of mobile devices. The platform also interfaces with nonmobile channels and devices such as e-mail, instant messenger, and Web. In an implementation, funds are accessed from an account holder's mobile device such as a mobile phone or a personal digital assistant to make or receive payments."
	Digital wallets	"A system and a method are provided for generating a digital receipt for purchases made utilizing a digital wallet or with other payment procedures. The digital receipt is stored in the cloud in a digital receipts repository for later retrieval. The digital receipt can be standardized to facilitate data processing of the data contained in data fields of the digital receipt."

TABLE II
PROPOSED FINTECH TAXONOMY, TOGETHER WITH APPLICATIONS CORRESPONDING TO EACH CATEGORIES, AND EXAMPLES OF PATENT ABSTRACTS.

classification task can be initialized with the parameters of a generic pre-trained BERT model, further pre-trained using FinTech patent data, and subsequently fine-tuned for patent classification. In the fine-tuning phase, the BERT architecture is similar to the architecture of a generic BERT model, except for the inclusion of a classification component (e.g., a fully connected layer, followed by a softmax classification layer). The classification component is linked to the first output embedding, C, corresponding to the first input token [CLS], and provides a representation of the whole input sequence. The input to the classification BERT model consists of tokens in a patent text (also preceded by [CLS]), and the output is the category of the input patent.

B. BERT Variants

The success of the initial BERT-based models has resulted in an unparalleled suite of variants that can be used with the pre-training/fine-tuning framework proposed in [5]. We used three variants in our analysis, specifically, RoBERTa [36], ALBERT [37] and XLNet [38]. RoBERTa (Robustly Optimized BERT Approach) [36] uses a larger dataset and an improved procedure to pre-train the BERT architecture. Among others, the next sentence prediction is removed and the masking applied to the training data is changed dynamically. ALBERT (A Lite BERT) [37] is focused on decreasing BERT's size (i.e., the number of parameters that need to be learned), while not hurting its performance. It achieves a reduction in the number of parameters by factorizing the embedding parameterization and sharing parameters across all layers. To improve the training, it replaces the next sentence prediction task with a sentence order prediction task that better captures the inter-sentence cohesion. XLNet [38] is a large bidirectional transformer, whose authors argue against the masked language modeling task and introduce an autoregressive permutation language modeling task for training (specifically, prediction of the next token in a sequence using some random order of the sequence). This improvement in the training procedure enables XLNet to capture better bidirectional dependencies among tokens in a sequence.

C. CNN Models

Convolutional Neural Networks (CNNs) [39], originally proposed in computer vision, have been successfully adapted to text classification [24]. In general, a CNN consists of convolutional layers followed by non-linear activations, pooling layers and fully connected layers. A convolutional layer employs a sliding window approach to apply a set of filters (low-dimensional tensors) to its input. The convolution operation captures local dependencies in the original input, and it produces feature maps. The pooling operation is used to reduce the dimensionality of the feature maps. Following convolutional layers (used together with non-linear activations), and pooling layers, a CNN has one or more fully connected layers, that capture non-linear dependencies among features [39]. The last fully connected layer uses a softmax activation function, and has as many output neurons as the number of targeted classes. In this work, we adopt the simple CNN architecture for text classification proposed in [24], which consists of one convolution layer with multiple filter widths and feature maps (and non-linear activations), followed by a max-pooling layer,

and finally a fully connected layer with softmax output. The input to our CNN models consists of vector embeddings of patent word tokens. We use a BERT model pre-trained on FinTech patent documents to represent input tokens.

D. RNN Models

LSTM (short for Long Short-Term Memory) networks [19] are a type of recurrent neural networks (RNN) that can be used to capture dependencies in sequence data, including long term dependencies. At the core of an RNN network, including LSTM networks, there is a recurrent cell represented as a hidden state, which enables the network to pass information from one time step to the next one. When unfolded, an RNN networks looks like a chain of repeated cells, which share the hidden state. General RNNs suffer from the gradient vanishing and exploding problem [19], when used with long sequences. LSTM networks avoid this problem by introducing a cell state, in addition to the internal hidden state, which carries information across the sequence. The information passed through the cell state is controlled by three gates, an *input gate*, a *forget gate* and an *output gate*. The gates help identify which information from the previous cell state needs to be forgotten, which information needs to be updated, and which information needs to be used as the output of the current cell state. A standard LSTM has one layer corresponding to one cell, and processes the sequence in the forward direction. A bidirectional LSTM (Bi-LSTM) includes a second layer/cell, which processes the sequence in reverse direction. In this work, we use a Bi-LSTM network to capture sequence dependencies in both forward and reverse directions. As input to the Bi-LSTM, we use BERT embeddings pre-trained on FinTech data.

V. Experimental Setup

In what follows, we formulate the research questions addressed in this work and describe the experiments that we design to answer these questions, and implementation details.

A. Research Questions

Our experimental setup is motivated by the following research questions:

(RQ1) When using different BERT-like models for FinTech patent classification, how do the results vary with the BERT model used? What model performs the best?

(RQ2) How do the results of the best BERT model compare with those of the CNN and RNN models?

(RQ3) How do the results of the deep learning models, including BERT, CNN and RNN models, compare with the results of the traditional machine learning approaches?

B. Experiments and Implementation Details

To answer the above research questions, we design and run the following experiments:

(RQ1) *BERT models.* Using the Transformers library by HuggingFace[4], we experiment with 22 pre-trained BERT models and variants, by fine-tuning the models using labeled data for our specific FinTech classification task. The models we experiment with, include BERT, RoBERTa and ALBERT. We use different architectures for each of these models (e.g., for BERT we used architectures such as: *bert-base-uncased*, *bert-base-cased*, *bert-large-uncased*, etc.). We train each model for 6 epochs, using the AdamW optimizer with a learning rate of $2e^{-5}$. We use default values for other hyper-parameters.

(RQ2) We compare the best BERT model for the FinTech patent classification task with models that train a CNN or an RNN on top of the patent pre-trained BERT model. The specific pre-trained model we use is 'bert-base-uncased' (for both the BERT tokenizer and BERT model). The pre-trained model creates embeddings of dimension 768 for each token in a patent abstract, and the token embeddings are fed as input to the CNN/RNN models. Any abstract that contains more than 512 tokens is truncated to exactly 512 tokens to accommodate the size limitation of BERT [5]. The weights of the BERT model are frozen during the CNN/RNN training.

Model Architecture for CNN. For the CNN network, we adopted the architecture proposed in [24]. The network has one convolution layer with multiple filter widths. Specifically, we used filters of width 3, 4, and 5, respectively. For each filter width, there are 32 feature maps, which are used to extract the features of a given patent abstract. The convolutions, are followed by a single layer feed-forward network that takes as input the flattened features extracted from the convolutions and outputs a probability distribution over classes. During training, we used the standard cross-entropy loss as our optimization criterion, and Adam as our optimizer. Further, we used a learning rate of $1e^{-4}$ that decayed by a factor of 10 every 5 epochs, for a total of 15 epochs. We found these particular hyper-parameters to work best by experimenting with a portion of the training subset as validation.

Model Architecture for RNN. The RNN architecture used corresponds to a Bi-LSTM network. The cells of the forward and reverse LSTMs that are part of the model have an input size of 768 to account for the BERT embeddings shape. Each LSTM has only one hidden layer with a size of 64. The model takes the element-wise max of the LSTM outputs and feeds that vector into a single layer fully connected network for classification. During training, we used the standard cross-entropy loss as our criterion and Adam as our optimizer. Further, we used a learning rate of $1e^{-4}$ over the course of 25 epochs. As in the case of the CNN models, we found these particular hyper-parameters to work best by experimenting with a portion of the training subset as validation.

(RQ3) *ML Baselines.* As traditional machine learning baselines, we experimented with machine learning approaches that were used in prior work on FinTech patent classification [4]. Specifically, we used Linear Support Vector Machines (L-SVM), Gaussian Support Vector Machines (G-SVM), Multilayer Perceptron (MLP), Naive Bayes (NB), Random Forest (RF), Gradient Boosting (GB), and also a voting classifier (ML-V). The voting classifier consists of L-SVM, G-SVM and MLP, similar to the voting classifier in [4], and uses the prediction made by the MLP, in case of ties. All the classifiers,

[4]https://github.com/huggingface/transformers

except for MLP, were implemented with `scikit-learn` library and used the default hyper-parameter values provided by the library (except for L-SVM, where $C = 0.0$, and G-SVM, where $C = 1$). The MLP network was implemented with PyTorch, and featured 4 layers. The first, second, and third layers had 256, 200, and 100 hidden neurons, respectively. During training, we used the standard cross-entropy as our optimization criterion and Adam as the optimizer. Further, we used a learning rate of $6e^{-4}$ for 30 epochs. One difference between our ML classifiers and the classifiers used in [4] is that we used patent pre-trained BERT embeddings for the features as opposed to the bag-of-words representation.

C. Evaluation Metrics

To evaluate the performance of the various models that we train, we use several standard metrics, including the overall accuracy, precision, recall and F1 scores. We also report the precision, recall, and F1 scores for each of our categories to determine what categories might be easier or harder to identify.

VI. Results and Discussion

In this section, we present the results of our experiments and discuss them in regard to the research questions raised.

(RQ1) *BERT models.* We fine-tune 22 pre-trained BERT-based models on our labeled training dataset, and estimate their performance in terms of F1-scores on the test dataset. A comparison of the 10 best-performing models is shown in Fig. 3. The F1-scores range from 89.39% (for "roberta-based") to 91.21% (for "bert-based-cased"), showing that the BERT-like models are generally good candidates for our FinTech patent classification problem. The best performing model, "bert-based-cased", has 12-layers, 768-hidden units, 12-heads and 110M parameters, and is pre-trained on cased English text. The next three best models (having similar scores to each other) are "roberta-large", "bert-large-cased" and "bert-large-cased-whole-word-masking", which all share the same architecture, specifically, 24-layers, 1024-hidden units, 16-heads, and 355M parameter, and are also trained on cased English text. As expected for patent documents, the uncased models have worse performance than the cased models overall.

Fig. 4 shows the confusion matrix corresponding to the best BERT model, which is later used for the large-scale classification of FinTech patents. The diagonal entries show the percentage of correctly classified instances in each category, while the non-diagonal entries show how the misclassified instances for a particular category are distributed among the other categories. As can be seen, the *Insurance*, *Investment*, *Fraud* and *Payments* have a high percentage of instances correctly classified (0.99%, 0.98%, 0.98% and 0.97%, respectively). For *Data Analytics*, 90% of instances are classified correctly, while 10% are misclassified as *Payments* (4%) or *Investment* (6%). Finally, the *Non-FinTech* category has 85% correctly classified instances, and the misclassified instances are spread across all the FinTech categories. Together, these results show that our models are effective at identifying true positives for the FinTech categories, although some false positives (Non-FinTech patents classified as FinTech) will also be expected.

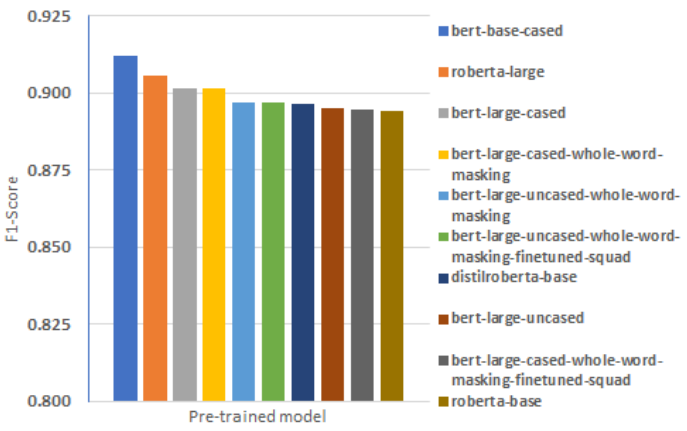

Fig. 3. Comparison of the 10 best-performing BERT-like models in terms of F1-Score. The `bert-base-cased` model has the best performance overall.

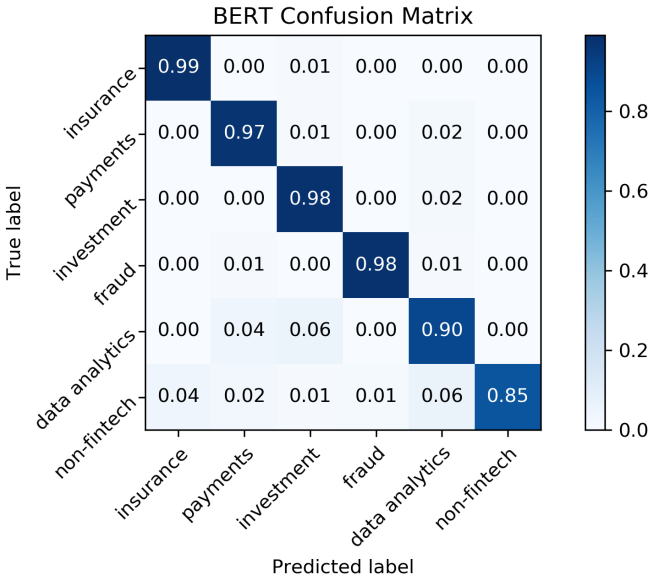

Fig. 4. Normalized confusion matrix corresponding to the best BERT model. The diagonal entries show the percentage of correctly classified instances in each category. Non-diagonal entries on a row show how the misclassified instances of a category are distributed among the other categories.

(RQ2) The result of the deep learning models, CNN and RNN, which take the patent pre-trained BERT embeddings as input, are shown on the right side of vertical double-line in Table III, by comparison with the results of the best BERT model fine-tuned on the patent training data. The table shows the results for each category separately, and also the average over the 6 categories captured by our labeled data (including 5 FinTech categories and 1 Non-FinTech category), and the accuracy of the models. The three models have similar results, although BERT is the best model overall, with an accuracy of 92.30% and an average F1-score of 91.20% over the six

categories. BERT is closely followed by the RNN model, while the CNN model has the worse performance among the three deep learning models. In terms of individual categories, the results show that the *Fraud* category has the best performance, with an F1-score of 97.00%, while the *Data Analytics* category has the worst performance, with an F1-score of 75.30%.

(RQ3) The results of the traditional machine learning models, which also take the patent pre-trained BERT embeddings as input are shown on the left side of the vertical double-line in Table III. As can be seen, the best model in this category is the MLP model. However, except for two cases when the RF model gives the best result in terms of precision for the *Data Analytics* category and the best result in terms of recall for the *Non-FinTech* category, respectively, the deep learning models are superior to the traditional machine learning models.

VII. Spatial Temporal Analysis

In this section, we provide a preliminary exploratory data analysis based on the FinTech dataset that has emerged after applying our best performing machine learning model (specifically, BERT). We also discuss both spatial and temporal trends across different types of FinTech subsectors.

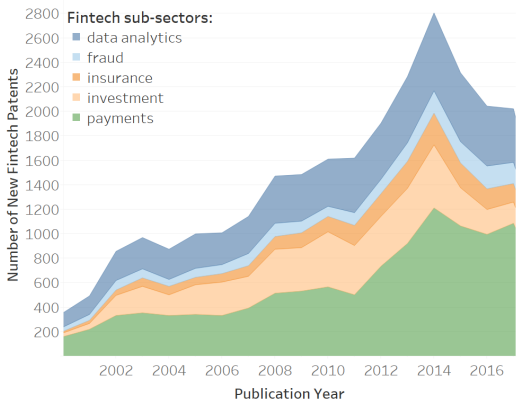

Fig. 5. FinTech innovation distribution over time (2000-2017), for the five FinTech categories in our taxonomy. FinTech growth peaks in the year 2014, and is driven by payment-related innovations, followed by investment-related and data analytics innovations.

Fig. 5 shows that FinTech innovations, proxied by patent applications to the US and EU patent offices, have grown sevenfold from just under 400 patents in 2000, to a peak of c. 2800 filings in the year 2014, and consistently delivering over 2000 innovations every year since. This growth has been fueled by payments-related innovations (e.g., digital wallets and low fee cross-country money transfer technologies), followed by investment and lending platforms, data analytics solutions, and to a lesser degree insurance and fraud detection tools. Our emerging distribution is largely congruent with the manual mapping/sampling of FinTech start-up innovations across the OECD by Cojoianu et al. [3], whose study, the only other one to our knowledge to provide cross-country evidence on FinTech innovations, is slightly biased towards lending and investment platforms, as this is the sub-sector that received the highest amount of venture capital to date. Our work improves on the study of [3], as it includes large scale evidence innovations coming from not only the start-up sector, but also financial and technology incumbents, as well as other entrants who innovate at the periphery of the sector.

As can be seen in Fig. 6, the distribution of FinTech innovations submitted to the US or EU patent offices shows the US taking the lead with c. 74% of the total innovations (19,361 patents), followed by Japan (5%), UK(4%), Germany, South and Canada (all three with 2%), and the rest of the countries with 1% or less. At the FinTech subsector level, the overall pattern across countries also seems to be prevalent within most countries, with payments, data analytics and investments/lending solutions representing the majority of innovations. France distinguishes here with a larger share of financial fraud detection innovations.

VIII. Conclusions and Future Work

We develop a FinTech taxonomy, and label a dataset according to this taxonomy to enable studies of FinTech innovations. We train BERT-based models to identify FinTech patents, and use them to shortlist a set of 25,580 FinTech patents in one of the following categories: *Data Analytics*, *Fraud*, *Insurance*, *Investments* and *Payments*. Subsequent temporal and spatial analysis of these patents shows a consistent growth in the FinTech sector, peaking in the year 2014, with the United States leading in terms of number of innovations.

As part of future work, we plan to further improve the models by using the whole text of the patent documents, as opposed to just the abstracts. We also plan to improve performance by using domain adaptation and transfer learning approaches, which can benefit from general patent data, in addition to FinTech data. The results of the spatial temporal analysis will be further used to gain understanding of what FinTech innovations are, where, when and why they emerge, and what their impact is on companies investing in them, and ultimately on the society. Finally, we believe that our taxonomy and dataset will help to substantially reduce the search costs for FinTech innovations, while also helping the financial sector and technology incumbents to understand the latest developments in FinTech.

References

[1] D. W. Arner, J. Barberis, and R. P. Buckey, "Fintech, regtech, and the reconceptualization of financial regulation," *Nw. J. Int'l L. & Bus.*, vol. 37, p. 371, 2016.

[2] D. Wójcik and T. F. Cojoianu, "A global overview from a geographical perspective," *Int. Financ. Centres after Glob. Financ. Cris. Brexit*, vol. 207, 2018.

[3] T. F. Cojoianu, G. L. Clark, A. G. Hoepner, V. Pazitka, and D. Wojcik, "Fin vs. tech: Determinants of fintech start-up emergence and innovation in the financial services incumbent sector," *Tech: Determinants of Fintech Start-Up Emergence and Innovation in the Financial Services Incumbent Sector*, 2019.

Category	Metric	L-SVM	G-SVM	NB	kNN	RF	GB	MLP	ML-V	CNN	RNN	BERT
Insurance	Pr	0.779	0.857	**0.878**	0.689	0.824	0.864	0.848	0.854	0.904	**0.922***	0.891
	Re	0.844	0.812	0.750	0.875	0.635	0.792	**0.927**	0.792	0.979	**0.990***	**0.990***
	F1	0.810	0.834	0.809	0.771	0.718	0.826	**0.886**	0.822	0.940	**0.955***	0.938
Payments	Pr	0.827	**0.852**	0.645	0.636	0.802	0.800	0.835	0.851	0.884	0.884	**0.906***
	Re	0.910	0.920	0.800	0.840	0.650	0.840	**0.960**	0.860	**0.990***	**0.990***	0.970
	F1	0.867	0.885	0.714	0.724	0.718	0.820	**0.893**	0.856	0.934	0.934	**0.937***
Investment	Pr	0.837	0.850	0.765	0.725	0.778	0.796	**0.870**	0.861	**0.943***	0.934	0.924
	Re	0.870	**0.960**	0.910	0.870	0.770	0.860	0.940	0.930	**0.990***	**0.990**	0.980
	F1	0.853	0.901	0.831	0.791	0.774	0.827	**0.904**	0.894	**0.966***	0.961	0.918
Fraud	Pr	0.788	0.790	0.676	0.617	0.813	0.815	0.849	**0.884**	0.951	**0.961***	0.960
	Re	0.780	**0.830**	0.750	0.740	0.610	0.750	0.790	0.760	**0.980***	**0.980***	**0.980***
	F1	0.784	0.810	0.711	0.673	0.697	0.781	**0.819**	0.817	0.966	**0.970***	**0.970***
Data Analytics	Pr	0.530	0.581	0.351	0.420	**0.700***	0.565	0.581	0.513	0.635	0.635	**0.652**
	Re	0.700	0.720	0.660	0.420	0.280	0.700	0.720	**0.800**	0.800	0.800	**0.900***
	F1	0.603	0.643	0.458	0.420	0.400	0.625	**0.643**	0.625	0.708	0.708	**0.753***
Non-FinTech	Pr	0.939	0.955	0.968	**0.980**	0.700	0.880	0.962	0.923	0.992	0.996	**1.000***
	Re	0.823	0.850	0.697	0.660	**0.927***	0.853	0.843	0.873	0.857	**0.863**	0.850
	F1	0.877	**0.899**	0.810	0.789	0.798	0.866	**0.899**	0.897	0.919	**0.925***	0.918
Average	Pr	0.783	0.814	0.714	0.678	0.770	0.787	**0.824**	0.814	0.885	**0.889***	**0.889***
	Re	0.821	0.849	0.761	0.734	0.645	0.799	**0.863**	0.836	0.933	0.935	**0.945***
	F1	0.799	0.829	0.722	0.695	0.684	0.791	**0.840**	0.819	0.905	0.909	**0.912***
	Acc	0.830	0.858	0.751	0.735	0.745	0.820	**0.867**	0.849	0.921	**0.924***	0.923

TABLE III

PERFORMANCE RESULTS ON THE TEST DATA FOR TRADITIONAL MACHINE LEARNING MODELS (ON THE LEFT OF THE VERTICAL DOUBLE-LINE) AND DEEP LEARNING MODELS (ON THE RIGHT OF THE LINE). THE PERFORMANCE IS REPORTED IN TERMS OF PRECISION (PR), RECALL (RE) AND F1-SCORE (F1), FOR EACH CATEGORY AND OVERALL. THE OVERALL ACCURACY (ACC) IS ALSO SHOWN AT THE END. THE BEST RESULTS FOR THE MACHINE LEARNING AND FOR THE DEEP LEARNING MODELS, RESPECTIVELY, ARE HIGHLIGHTED IN BOLDFACE. THE BEST RESULTS OVERALL IN A ROW ARE ALSO MARKED WITH *. THE BEST MACHINE LEARNING MODEL IS THE MLP MODEL, WHILE THE BEST DEEP LEARNING MODEL IS THE BERT MODEL.

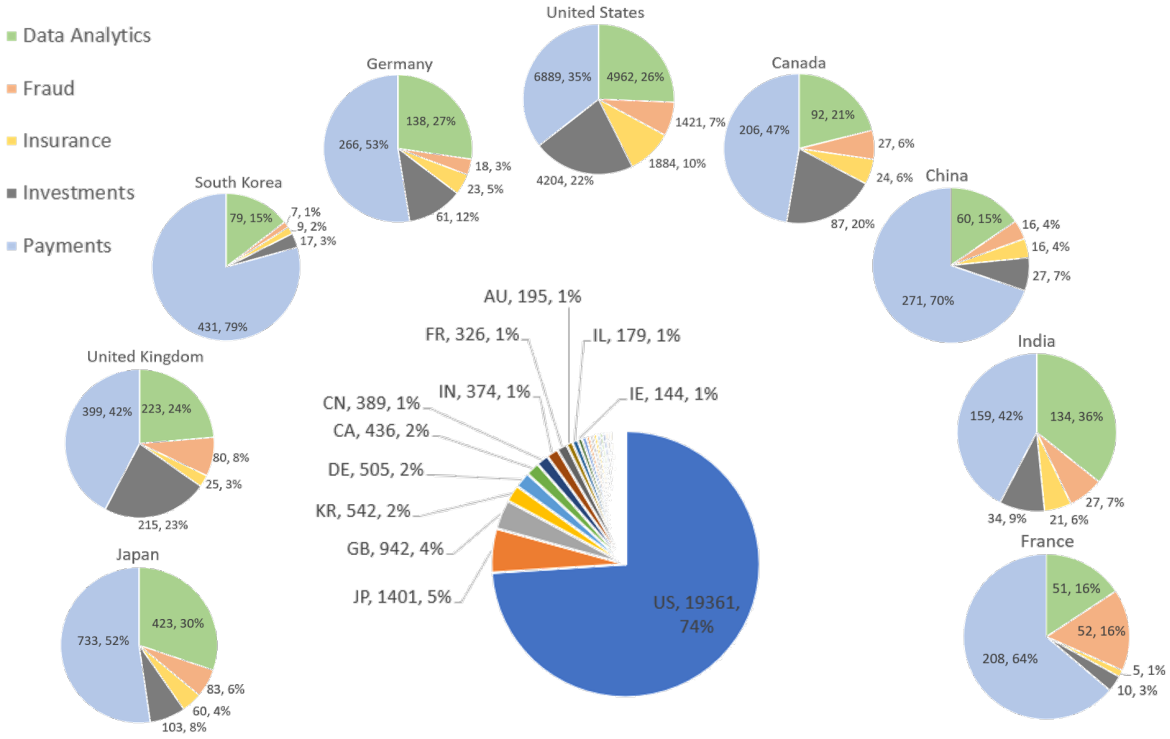

Fig. 6. Cross-country cumulative FinTech innovation distribution over 2000-2017 for the five FinTech innovation categories. US is leading with over 74% of total innovations filed to the EPO and USPTO, followed by Japan, UK and Germany. Top 9 countries with FinTech innovations have an uneven distribution across different technologies, but payments-related innovations feature most in each country.

[4] M. A. Chen, Q. Wu, and B. Yang, "How valuable is fintech innovation?" *The Review of Financial Studies*, vol. 32, no. 5, pp. 2062–2106, 2019.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[6] P. Gomber, J.-A. Koch, and M. Siering, "Digital finance and fintech: current research and future research directions," *Journal of Business Economics*, vol. 87, no. 5, pp. 537–580, 2017.

[7] I. Goldstein, W. Jiang, and G. A. Karolyi, "To fintech and beyond," *The Review of Financial Studies*, vol. 32, no. 5, pp. 1647–1661, 2019.

[8] P. Gomber, R. J. Kauffman, C. Parker, and B. W. Weber, "Financial information systems and the fintech revolution," 2018.

[9] C. J. Fall, A. Törcsvári, K. Benzineb, and G. Karetka, "Automated categorization in the international patent classification," in *Acm Sigir Forum*, vol. 37, no. 1. ACM New York, NY, USA, 2003, pp. 10–25.

[10] K. Benzineb and J. Guyot, "Automated patent classification," in *Current challenges in patent information retrieval*. Springer, 2011, pp. 239–261.

[11] J. C. Gomez and M.-F. Moens, "A survey of automated hierarchical classification of patents," in *Prof. search in the modern world*. Springer, 2014, pp. 215–249.

[12] M. F. Grawe, C. A. Martins, and A. G. Bonfante, "Automated patent classification using word embedding," in *16th IEEE Int. Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017, pp. 408–411.

[13] S. Li, J. Hu, Y. Cui, and J. Hu, "Deeppatent: patent classification with convolutional neural networks and word embedding," *Scientometrics*, vol. 117, no. 2, pp. 721–744, 2018.

[14] M. Shalaby, J. Stutzki, M. Schubert, and S. Günnemann, "An lstm approach to patent classification based on fixed hierarchy vectors," in *Proceedings of the 2018 SIAM Int. Conference on Data Mining*. SIAM, 2018, pp. 495–503.

[15] J. Hu, S. Li, J. Hu, and G. Yang, "A hierarchical feature extraction model for multi-label mechanical patent classification," *Sustainability*, vol. 10, no. 1, p. 219, 2018.

[16] L. Abdelgawad, P. Kluegl, E. Genc, S. Falkner, and F. Hutter, "Optimizing neural networks for patent classification," in *Proc. of ECML-PKDD*, vol. 16, 2019.

[17] J.-S. Lee and J. Hsiang, "Patentbert: Patent classification with fine-tuning a pre-trained bert model," *arXiv preprint arXiv:1906.02124*, 2019.

[18] L. Xu, X. Lu, G. Yang, and B. Shi, "Identifying fintech innovations with patent data: A combination of textual analysis and machine-learning techniques," in *Int. Conference on Information*. Springer, 2020, pp. 835–843.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comp.*, vol. 9, no. 8, 1997.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[21] J. Risch and R. Krestel, "Domain-specific word embeddings for patent classification," *Data Technologies and Applications*, 2019.

[22] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. of the ACL*, vol. 5, pp. 135–146, 2017.

[23] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *EMNLP*, vol. 14, 2014, pp. 1532–1543.

[24] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[25] S. Chae and J. Gim, "A study on trend analysis of applicants based on patent classification systems," *Information*, vol. 10, p. 364, 2019.

[26] M. Sofean, H. Aras, and A. Alrifai, "Analyzing trending technological areas of patents," in *Int. Conf. on DEXA*. Springer, 2019, pp. 141–146.

[27] B. Mellon, "Innovation in payments: The future is fintech," *The Bank of Newyork*, 2015.

[28] R. Levy, "Fintech market map," *Retrieved*, vol. 11, no. 08, p. 2016, 2015.

[29] F. Amalia, "The fintech book: The financial technology handbook for investors, entrepreneurs and visionaries," *Journal of Indonesian Economy and Business*, vol. 31, no. 3, pp. 345–348, 2016.

[30] E. . Young and G. B. Treasury, *UK Fintech on the Cutting Edge: An Evaluation of the International Fintech Sector*. Ernst & Young.

[31] CBInsights, "Wealth tech market map," 2017.

[32] J. Eckenrode and S. Friedman, "Fintech by the numbers," *Deloitte Services LP.*, 2017.

[33] C. Haddad and L. Hornuf, "The emergence of the global fintech market: Economic and technological determinants," *Small Business Economics*, vol. 53, no. 1, pp. 81–105, 2019.

[34] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?" in *Nat. Conf. on Chinese Comp. Linguistics*. Springer, 2019, pp. 194–206.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[36] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[37] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," in *ICLR*, 2020.

[38] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," in *Advances in NIPS*, 2019, pp. 5754–5764.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.