



## Análise de Viabilidade

### **ESTUDO DAS FERRAMENTAS A SEREM UTILIZADAS NA TRANSCRIÇÃO DOS ÁUDIOS**

## HISTÓRICO

<b>Data</b>	<b>Versão</b>	<b>Descrição</b>	<b>Autor</b>
<b>16/09/2019</b>	1.0	Estudo das soluções de mercado a serem usadas na transcrição dos áudios.	Marcello S. Pinheiro
<b>30/09/2019</b>	1.1	Revisão e finalização.	Marcello S. Pinheiro
<b>07/10/2019</b>	1.2	Inclusão da proposta de desenvolvimento da arquitetura Deep Learning para Transcrição.	Marcello S. Pinheiro
<b>10/10/2019</b>	2.0	Ajustes e correções.	Marcello S. Pinheiro

## ÍNDICE

INTRODUÇÃO.....	4
1. CONTEXTUALIZAÇÃO.....	5
2. SOLUÇÕES DE MERCADO.....	5
2.1. GOOGLE SPEECH.....	5
2.2. MICROSOFT VIDEO INDEXER.....	6
2.3. AMAZON TRANSCRIBE.....	7
2.4. IBM WATSON API.....	8
2.5. MICROSOFT SPEECH.....	9
3. TABELA COMPARATIVA DE CUSTO ENTRE AS API SPEECH-TO-TEXT.....	10
4. CUSTO DAS SOLUÇÕES STANDARD DE MERCADO X MINUTOS.....	11
5. PROPOSTA DE DESENVOLVIMENTO DE UMA ARQUITETURA DEEP LEARNING.....	14
CONCLUSÃO.....	16

## INTRODUÇÃO

Atualmente, a Inteligência Artificial (IA) tem o potencial de contribuir, fortemente, na simplificação e na otimização das atividades diárias. No entanto, diversos indivíduos, empresas e até desenvolvedores resistem em adotar essa tecnologia. Tal resistência está relacionada com a curva de aprendizado das tecnologias de IA, sendo transposta através da utilização de outras empresas que já prestam tais serviços de forma consolidada.

Uma dessas tecnologias é a transcrição de arquivos no formato áudio para arquivos texto, do termo Speech-to-Text, muito usada como solução de problemas de negócio, por exemplo, quando se deseja gravar e consolidar o registro gravado de reuniões até volumes gigantescos de dados em centrais de atendimento telefônico.

Nesse documento serão feitas comparações de cinco grandes empresas que prestam esse tipo de serviço nas dimensões: precificação, características e qualidade.

Também será apresentada uma proposta de desenvolvimento de uma arquitetura Deep Learning para as transcrições, a qual será comparada, em termos de custo, com as soluções de mercado.

## 1. CONTEXTUALIZAÇÃO

Nesse primeiro momento, o presente estudo mostra uma comparação entre precisificação, características e qualidade das soluções de mercado para a transcrição de áudios em textos. No entanto, também servirá como parâmetro para decidir e justificar o desenvolvimento de uma Arquitetura Deep Learning específica para o discurso dos áudios dos atendimentos gravados na Central de Atendimento da Coordenação de Relacionamento da Ouvidoria Geral do Ministério da Cidadania.

O estudo que norteou o presente documento consistiu em explorar o que as empresas de tecnologia em atuação no mercado atualmente estão usando de Inteligência Artificial (IA) para a transcrição de arquivos em formato áudio para arquivos texto. Desse modo, muitas são as empresas com diferentes portes e especificidades que apresentam soluções Speech-to-Text. A pesquisa se voltou para as empresas mais conhecidas de grande porte que prestam serviços de tecnologia oferecendo essas soluções de IA.

As empresas selecionadas foram: Google, Amazon, IBM e Microsoft. Tais empresas possuem soluções que funcionam como um tipo de assistente pessoal com inteligência para transcrever áudios em textos.

Em suas operações diárias, essas empresas estendem suas soluções Speech-to-Text a indivíduos para serem testadas gratuitamente. Essas ofertas variam de tarefas como fala para texto e melhor funcionalidade de pesquisa, até análises de vídeo extremamente complexas, que envolvem reconhecimento facial e identificação do orador.

As soluções de mercado estudadas foram:

1. Google Cloud Speech API;
2. Amazon Transcribe Web Services;
3. IBM Watson Speech-toText API;
4. Microsoft Cognitive Services Speech-to-Text;
5. Microsoft Cognitive Services Video Indexer.

## 2. SOLUÇÕES DE MERCADO

### 2.1. Google Speech

A Google possui uma API, encontrada em (23/09/2019), oferecida no advento nuvem, que preenche um viés entre usabilidade *versus* funcionalidade.

Como qualquer tecnologia, no início, essa API tem uma curva de aprendizado, mas oferece uma boa documentação e com exemplos que contribuem para acelerar a aprendizagem.

Tal documentação permite que um usuário sem experiência programe por meio da CLI (Command Line Interface).

Abaixo segue a tabela 1 que mostra a relação benefício x desvantagem da API da Google:

BENEFÍCIOS	DESVANTAGENS
Disponibiliza um crédito de 300 dólares para usar na nuvem da Google livremente.	É a segunda pior precisão (qualidade) de transcrição.
É uma das empresas que possui grande eficiência (velocidade) na transcrição dos arquivos.	Tem o custo de USD\$ 0.006/15s em transcrições de 60 minutos a 1 milhão de minutos na versão padrão.
Não requer conhecimento de programação para usar a API por meio da CLI, a qual possibilita desenvolver de forma eficiente.	A atualização que aumenta a precisão da transcrição dos áudios dobra o custo para USD\$ 0.012/15s.
Inclui recursos e extensões que permitem aumentar a flexibilidade e qualidade.	

Tabela 1 – API Google – Benefícios x Desvantagens

## **2.2. Microsoft Video Indexer**

Trata-se de uma das APIs mais poderosas e é encontrada em <https://azure.microsoft.com/en-us/services/media-services/video-indexer/> (23/09/2019). Oferece um portal web intuitivo, além de uma documentação boa para programadores.

Um aspecto a favor do Video Indexer é a possibilidade de usar uma ampla variedade de soluções de IA e possui recursos adicionais para desenvolvedores como, por exemplo, treinar modelos personalizados.

Possui a melhor avaliação gratuita, em relação aos demais serviços, e suporta vários tipos de arquivos. Todos esses recursos tornam o Video Indexer um dos serviços mais fáceis e baratos, em relação a custo x benefício, de usar sem, com isso, comprometer qualquer precisão ou flexibilidade.

Abaixo segue a tabela 2 que mostra a relação benefício x desvantagem da API do Microsoft Video Indexer:

BENEFÍCIOS	DESVANTAGENS
É a segunda melhor em precisão.	É o segundo maior tempo de processamento de arquivos.
API mais fácil de aprender e de usar.	O custo para usar toda a gama de serviços pode ficar dispendioso. O standard custa USD\$ 0.015/min e no premium USD\$ 0.035/15min.
Avaliação gratuita permite 100 vídeos/áudios enviados diariamente.	
Melhor relação custo x benefício.	
Tem o maior pool de tipos de arquivos suportados.	
Dá a possibilidade de treinar modelos personalizados.	

Tabela 2 – API Microsoft Video Indexer – Benefícios x Desvantagens

### **2.3. Amazon Transcribe**

Trata-se de uma das várias ferramentas de machine learning que a Amazon oferece em seus serviços AWS. Oferece uma interface web fácil de usar por qualquer pessoa no console da AWS.

No geral, o Amazon Transcribe, <https://aws.amazon.com/pt/transcribe/> (23/09/2019), é fácil de usar e possui uma das transcrições mais precisas do mercado. Em contrapartida, é o custo mais alto de todos os demais porque enfatiza a mentalidade de receber pelo que é pago.

Abaixo segue a tabela 3 que mostra a relação benefício x desvantagem da API do Amazon Transcribe:

BENEFÍCIOS	DESVANTAGENS
É a transcrição mais precisa dentre os demais.	Custo de USD\$ 0.006/15s
Tem uma boa usabilidade.	É o maior tempo de processamento de arquivos.
Segundo maior pool de tipos de arquivos de áudios suportados.	
No primeiro ano de uso a versão <i>free</i> disponibiliza os primeiros 60 minutos de transcrição de áudio gratuitos por meses.	
Um dos que possuem a tarefa de personalizar uma lista de palavras.	

Tabela 3 – API Amazon Transcribe – Benefícios x Desvantagens

## **2.4. IBM Watson API**

A suíte Watson, <https://www.ibm.com/watson/services/speech-to-text/> (23/09/2019) possui uma boa API de desenvolvimento para programadores e, por outro lado, exige mais estudo do usuário leigo.

O tempo que é gasto antes de fazer funcionar algum modelo de transcrição é justificado pela eficiência do serviço.

Apesar do tempo de aprendizado levado para aprender a API, é o mais eficiente em processamento.

Em outras palavras, a API Watson oferece o melhor custo x benefício entre velocidade e precisão.

Abaixo segue a tabela 4 que mostra a relação benefício x desvantagem da API Watson IBM:



BENEFÍCIOS	DESVANTAGENS
É um dos melhores em precisão.	Tradicionalmente os primeiros 250.000 minutos tem o custo de USD\$ 0.02/min ou USD\$ 0.005/15seg.
Possui a melhor precisão <i>versus</i> relação de velocidade. É o mais rápido transcritor.	Requer conhecimentos e habilidades de programação.
Plano lite tem 500 minutos de transcrições gratuitas por mês.	
O custo do serviço diminui a partir de 250.000 minutos de áudio por mês. Pode chegar a 0.60/h ou 0.00016/s dólares se forem transcritos mais de 1.000.000 de minutos de áudio ou aproximadamente 2 anos em áudio/vídeo.	

Tabela 4 – API Watson IBM – Benefícios x Desvantagens

## **2.5. Microsoft Speech**

Essa API ficou aquém em relação aos serviços anteriores, apesar de ter a melhor velocidade de processamento. A precisão da transcrição é baixa e a documentação carece de atualização.

Os problemas de documentação praticamente inviabilizam o uso aos que não possuem conhecimento de programação em SDK.

Embora as transcrições longas sejam imprecisas, tem, em contrapartida, uma API que fornece uma solução rápida para o reconhecimento de voz.

Em linhas gerais, possui transcrições curtas e úteis e os preços mais módicos e de uso prolongado. Entretanto, a curva de aprendizado é alta.

Abaixo segue a tabela 5 que mostra a relação benefício x desvantagem da API Microsoft Speech-to-Text:

BENEFÍCIOS	DESVANTAGENS
Possui a maior velocidade de processamento de arquivos.	A avaliação gratuita é de 5h por mês para teste.
Tem o serviço mais barato a USD\$ 1/h ou USD\$ 0.00025/15s.	Tem a precisão de transcrição ruim.
Tradução e lista de palavras personalizadas oferecidas que podem melhorar a precisão dos comandos de fala.	A documentação do SDK é desatualizada.
	Exige conhecimentos e habilidade de programação.

Tabela 5 – API Microsoft Speech-to-Text – Benefícios x Desvantagens

### 3. TABELA COMPARATIVA DE CUSTO ENTRE AS API SPEECH-TO-TEXT

DIMENSÃO		AWS Transcribe Standard	Google Speech Standard	Microsoft Video Indexer Standard	IBM Speech API Lite	Microsoft Speech Standard
PRECIFICAÇÃO USD\$		0.006/15s	0.006/15s	0.0087/15s	0.005/15seg	0.00025/15s
CARACTERÍSTICAS	Identificação da Fala	X	X	X	X	
	Lista de Palavras Customizadas		X	X	X	X
	Deteção de Linguagem	X	X			
QUALIDADE Precisão mundial classificada com a ferramenta NIST Sclite NIST SCKT		49%	32%	48%	47%	18%

Tabela 6 – Comparação de preço, característica e qualidade.

#### 4. CUSTO DAS SOLUÇÕES STANDARD DE MERCADO x MINUTOS

A tabela 7 abaixo foi montada no intuito de projetar o custo das soluções standard de mercado *versus* a quantidade de arquivos processados por mês.

Na primeira coluna consta a projeção de 01 minuto a 05 minutos de arquivos áudio transcritos para arquivos texto. Ainda, nessa coluna tem a conversão de minutos para a equivalência em 15 segundos, haja vista que a precificação da tabela 6 acima está estipulada em fatores de 15 segundos permitindo, dessa forma, a comparação de preços entre as soluções. As demais colunas são as precificações de cada uma das cinco soluções de mercado estudadas.

Por exemplo, a solução **Google Speech** possui o custo de **USD\$ 0.006 dólares** para cada **15 segundos** de transcrições. Seriam **0.006 x 4** porque **01 minuto** contém **4 ¼ de 15 segundos**, **2 minutos** contém **8 ¼ de 15 segundos** e assim sucessivamente.

O resultado de **4 x 0.006** é igual a **0.024** dólares. Se for considerado **1000** (mil) arquivos transcritos diários seria **0.024 x 1000** dando **24 dólares diários**. Multiplicando **24 dólares** diários por **22 dias úteis** dá um **custo mensal** de **528 dólares**. Tudo isso se for transcrito, somente, 01 minuto de áudio para cada 1000 atendimentos diários por dia útil no mês.

Minutos Segundos	AWS Transcribe	Google Speech	Microsoft Video Indexer	IBM Speech API	Microsoft Speech
	Custo em USD\$ por 15 segundos				
	0.006	0.006	0.0087	0.005	0.00025
1m 4 x 15seg	4 x 0.006 = 0.024 x 1000 = 24 x 22 = <b>528</b>	4 x 0.006 = 0.024 x 1000 = 24 x 22 = <b>528</b>	4 x 0.0087 = 0.0348 x 1000 = 34.8 x 22 = <b>765.60</b>	4 x 0.005 = 0.02 x 1000 = 20 x 22 = <b>440</b>	4 x 0.00025 = 0.001 x 1000 = 1 x 22 = <b>22</b>
2m 8 x 15seg	8 x 0.006 = 0.048 x 1000 = 48 x 22 = <b>1,056</b>	8 x 0.006 = 0.048 x 1000 = 48 x 22 = <b>1,056</b>	8 x 0.0087 = 0.064 x 1000 = 64 x 22 = <b>1,408</b>	8 x 0.005 = 0.04 x 1000 = 40 x 22 = <b>880</b>	8 x 0.00025 = 0.002 x 1000 = 2 x 22 = <b>44</b>

Minutos Segundos	AWS Transcribe	Google Speech	Microsoft Video Indexer	IBM Speech API	Microsoft Speech
	Custo em USD\$ por 15 segundos				
	0.006	0.006	0.0087	0.005	0.00025
3m 12 x 15seg	$12 \times 0.006 =$ $0.072 \times 1000 =$ $72 \times 22 = \mathbf{1,584}$	$012 \times 0.006 = .$ $072 \times 1000 = 72$ $\times 22 = \mathbf{1,584}$	$12 \times 0.0087 =$ $0.1044 \times 1000 =$ $104.40 \times 22 =$ <b>2,296.8</b>	$12 \times 0.005 =$ $0.06 \times 1000 = 60$ $\times 22 = \mathbf{1,320}$	$12 \times 0.00025 =$ $0.003 \times 1000 = 3$ $\times 22 = \mathbf{66}$
4m 16 x 15seg	$16 \times 0.006 =$ $0.096 \times 1000 =$ $96 \times 22 = \mathbf{2,112}$	$16 \times 0.006 =$ $0.096 \times 1000 =$ $96 \times 22 = \mathbf{2,112}$	$16 \times 0.0086 =$ $0.128 \times 1000 =$ $128 \times 22 = \mathbf{2,816}$	$16 \times 0.005 =$ $0.08 \times 1000 = 80$ $\times 22 = \mathbf{1,760}$	$16 \times 0.00025 =$ $0.004 \times 1000 = 4$ $\times 22 = \mathbf{88}$
5m 20 x 15seg	$20 \times 0.006 =$ $0.12 \times 1000 =$ $120 \times 22 = \mathbf{2,640}$	$20 \times 0.006 =$ $0.12 \times 1000 =$ $120 \times 22 = \mathbf{2,640}$	$20 \times 0.0087 =$ $0.16 \times 1000 =$ $160 \times 22 = \mathbf{3,520}$	$20 \times 0.005 =$ $0.10 \times 1000 =$ $100 \times 22 = \mathbf{2,200}$	$20 \times 0.00025 =$ $0.006 \times 1000 = 6$ $\times 22 = \mathbf{132}$

Tabela 7 – Custo mensal em dólares para cada 1000 arquivos transcritos diários por mês.

Se for considerada uma média de 04 minutos de transcrições para cada áudio registrado, tempo esse razoável para conter os termos mais usados no discurso dos atendimentos do Projeto Fala Cidadão, o custo mensal das soluções de mercado para cada 1000 arquivos diários por mês seria:

1. Microsoft Video Indexer: USD\$ 2,816 dólares por mês;
2. Amazon Transcribe: USD\$ 2,640 dólares por mês;
3. Google Speech: USD\$ 2,640 dólares por mês;
4. IBM Speech API: USD\$ 1,770 dólares por mês;
5. Microsoft Speech: USD\$ 132 dólares por mês;

Apesar do custo da solução Microsoft Speech ser a mais barata, em contrapartida é uma solução cuja curva de aprendizado é bem longa devido à complexidade de sua arquitetura, é a de menor precisão e mais lenta em processamento, além da documentação e do framework estarem desatualizados em relação às demais soluções estudadas.

A solução Microsoft Video Indexer é a de custo mais alto, porém, é a de melhor precisão, contém uma das melhores documentações, é a terceira mais rápida em processamento e possui recursos extras de Inteligência Artificial já inclusos nesse preço, ou seja, não há custo a mais caso queiram os desenvolvedores customizarem e

retreinarem o modelo de transcrição. Obviamente, para isso é requerido um conhecimento mais profundo dessa solução, bem como o entendimento dos diversos tipos de métodos e técnicas/algoritmos disponíveis.

As soluções Amazon Transcribe e Google Speech possuem o mesmo custo mensal, porém, a solução Amazon Transcribe possui a terceira melhor precisão e a solução Google Speech é a quarta de melhor precisão ganhando em termos de precisão, somente, da Microsoft Speech e é a quarta mais lenta entre todas.

A solução **IBM Speech API** é a de **melhor custo x benefício** posto que, depois de aprendida a usabilidade é a segunda de melhor precisão, perdendo somente para o Microsoft Video Indexer, é a mais rápida em termos de processamento dos áudios e a segunda mais barata em termos de custo.

Cada uma das cinco soluções gera o seguinte **custo anual**, considerando 04 minutos de áudios transcritos para cada 1000 atendimentos diários por mês:

1. Microsoft Video Indexer: USD\$ 33,792 dólares por ano;
2. Amazon Transcribe: USD\$ 31,680 dólares por ano;
3. Google Speech: USD\$ 31,680 dólares por ano;
4. IBM Speech API: USD\$ 21,240 dólares por ano;
5. Microsoft Speech: USD\$ 1,584 dólares por ano;

## 5. PROPOSTA DE DESENVOLVIMENTO DE UMA ARQUITETURA DEEP LEARNING

O desenvolvimento de arquiteturas Deep Learning (DL) feitas na linguagem Python explicadas em [1], [2], [3], [6], [9], [10], [13] e [15] embasam-se nas *library* TensorFlow, Keras e Scikit-learn, sendo as duas primeiras as mais utilizadas devido o suporte ao desenvolvimento sobre GPU (Graphics Processing Unit), além do suporte às CPU (Central Processing Unit). Elas possuem um vasto recurso para arquiteturas DL como Long Short-Term Memory (LSTM) e Convolutional Neural Networks (CNN), amplamente utilizadas em situações-problema de reconhecimento de imagens (Image Recognition) e em transcrições de áudio para texto (Speech-to-Text).

Nas referências [3], [6], [9], [10] e [13] são discutidas as arquiteturas DL para Speech-to-Text. Nelas são explicados que a LSTM é a DL mais apropriada para o problema em questão do projeto Fala Cidadão. Além disso, há um amplo suporte de documentação e esclarecimentos quanto ao desenvolvimento dessa solução para rodar nas GPU. Em outras palavras, existe um amplo desenvolvimento de soluções Speech-to-Text que usam a arquitetura LSTM, vários casos de sucesso e reconhecidamente útil para resolver a situação-problema em questão.

A solução Speech-to-Text do projeto Fala Cidadão requer uma LSTM:

- Que possibilite o processamento em GPU, devido ao gigante volume de áudios;
- Que seja customizada aos discursos entre os atendentes e os cidadãos que recebem auxílio governamental, a fim de obter melhor acurácia e precisão nas transcrições de áudio para texto;
  - A utilização dos termos pertinentes ao negócio na fase de treinamento dessa arquitetura possibilitará uma grande vantagem em termos de baixo custo de desenvolvimento, alta acurácia, precisão e independência das soluções de mercado.
- Que possibilite o retreinamento e o desenvolvimento contínuo a custo mínimo de manutenção;
- Que tenha um custo de desenvolvimento abaixo das soluções de mercado.

As soluções DL que usufruem da arquitetura computacional Spark/Hadoop, como mostrado em NANDI [7] e em SHERIF e RAVINDRA [8], mostram o potencial de tal arquitetura, mas não é considerada a utilização das GPU no processamento. Tal tecnologia, a que mescla Spark/Hadoop com processamento em conjunto com as GPU é recente, carece de mais documentação explicativa e requerem mais pesquisa e aprofundamento para a sua devida utilização. Nesse momento essa solução não é considerada por conta do curto prazo de desenvolvimento do presente projeto.

Estimam-se 40 dias úteis de desenvolvimento, mais 5 dias úteis para verificação da qualidade, ajuste e otimização da acurácia e precisão da transcrição dos arquivos em formato áudio para arquivos textos, totalizando 45 dias úteis.

Os primeiros 40 dias úteis de desenvolvimento serão divididos em 15 dias de estudos para aprofundar e embasar a arquitetura proposta, e mais 30 dias para a implementação dessa solução, incluindo o primeiro nível de treinamento da arquitetura DL LSTM. O desenvolvimento terá um custo total de XXX (xxxxx). Entretanto, esse custo é somente inicial e não ocorrerá mensalmente. Posteriormente, sendo o caso e sob demanda, pode haver ajustes, adequação e retreinamentos ao custo da UST de desenvolvimento.

## CONCLUSÃO

Os preços dos serviços das soluções ora analisadas foram pesquisados colocando o termo-chave “*custo <nome-empresa> speech por segundo*”. Como a primeira solução pesquisada foi a da Google, que utiliza a precificação em valor por 15 segundos, foi mantida essa escala a fim de possibilitar a comparação no custo dos serviços entre as demais soluções de mercado estudadas. Todos os preços aferidos foram baseados nas soluções standard.

O produto da IBM apresentou a melhor relação custo x benefício, pois além de possuir o segundo melhor preço é o de melhor precisão e um dos mais rápidos em termos de processamento.

A solução Vídeo Indexer da Microsoft possui o melhor conjunto de IA e um dos melhores em Speech-to-Text, além de possibilitar a vantagem de treinar o modelo. Porém, é o mais caro dentre todos.

As soluções AWS Transcribe e Google Speech são medianas em relação a preço, precisão e desempenho.

Por último, a solução Microsoft Speech que apesar de possuir o menor custo tem, em contrapartida, API mais complexa que requer bastante experiência em programação e conhecimento dos métodos e técnicas/algoritmos desse framework.

A proposta de uma arquitetura Deep Learning Speech-to-Text desenvolvida em Python poderia atender plenamente ao problema de negócio do Projeto Fala Cidadão sem custo mensal de utilização, apenas o custo inicial de desenvolvimento. Essa solução terá um custo inicial para ser desenvolvido um transcritor específico e mais adequado ao discurso do projeto Fala Cidadão. Pode ser construído, treinado e otimizado conforme os requerimentos e necessidades do negócio em questão. Além disso, dá a possibilidade de retreinar sempre que for preciso a fim de melhorar a precisão e acurácia, e permite utilizar os descritores mais apropriados e inerentes ao negócio na hora que for necessário. Com isso, as chances de atingir a melhor precisão, acurácia e desempenho aumentam significativamente em função de ser uma solução totalmente personalizada e customizada para o negócio Fala Cidadão.



## REFERÊNCIAS

- [1] BENGFORT, Benjamin, BILBRO, Rebecca and OJEDA, Tony. Applied Text Analysis with Python. O'Reilly, 2018.
- [2] BEYSOLOW II, Taweh. Applied Natural Language Processing with Python. Apress, 2018.
- [3] LYTICA. Speech-to-Text Recognition. medium.com consultado em 10/09/2019.
- [4] SRINIVASA-DESIKAN, Bhargav. A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Packt Publishing, 2018.
- [5] PERKINS, Jacob. Python 3 Text Processing with NLTK 3 Cookbook. Packt Publishing, ed. 2, 2014.
- [6] BERNICO, Mike. Deep Learning Quick Reference. Packt Publishing, 2018.
- [7] NANDI, Amit. Spark for Python Developers. Packt Publishing, 2015.
- [8] SHERIF, Ahmed, RAVINDRA, Amrith. Apache Spark Deep Learning. Packt Publishing, 2018.
- [9] RAMSUNDAR, Bharath, ZADEH, Reza Bosagh. TensorFlow for Deep Learning: From linear regression to reinforcement learning. O'Reilly, 2018.
- [10] DI, Wei, BHARDWAJ, Anurag and WEI, Jianing. Deep Learning Essentials. Packt Publishing, 2018.
- [11] ALBON, Chris. Machine Learning with Python Cookbook. O'Reilly, 2018.
- [12] DAS, Sibanje. CAKMAK, Umit Mert. Hands-On Automated Machine Learning. Packt Publishing, 2018.
- [13] MANASWI, Navin Kumar. Deep Learning with Applications Using Python. Apress, 2018.
- [14] ANKAN, Ankur. PANDA, Abinash. Hands-On Markov Models with Python. Packt Publishing, 2018.
- [15] ZACCONE, Giancarlo, KARIM, Md. Rezaul. Deep Learning with TensorFlow, 2ed. Packt Publishing, 2018.