EXECUTIVE SUMMARY

Veritext harnesses the power of machine learning to classify large pieces of text into its respective category or categories saving legal executives time and effort. Hundreds of thousands of pieces of text of up to but not limited to 5000 words can be classified within seconds with acceptable accuracy considering the infancy of the product.

Future prospects include developing state-of-the-art deep learning algorithms to classify the texts, incorporating an 'Online Learning' feature so the algorithm can learn from new data in real time, using more advanced machine learning algorithms to summarise long documents and including AI-driven chatbots to assist users with their queries.

1.1. The Model

Machine Learning is a novel approach to automating complex problems with a long list of elaborate rules. Instead of explicitly telling the computer what to do and hardwiring every rule, data is fed to the computer with appropriate identifiers. Minimising a loss function, the computer is guided toward the rules in a hot/cold method. This minimization of the loss function is attributed to the 'learning' ability of a computer. This is an example of supervised learning.

Machine Learning can be used for a never-ending list of problems including various types of Natural Language Processing (NLP). NLP is the process of wrangling natural text used by humans and converting it into a more computer friendly format which retains the meaning of the text. This mined text can then be used for various purposes such as exploring the common words around a theme, or even a plethora of machine learning uses.

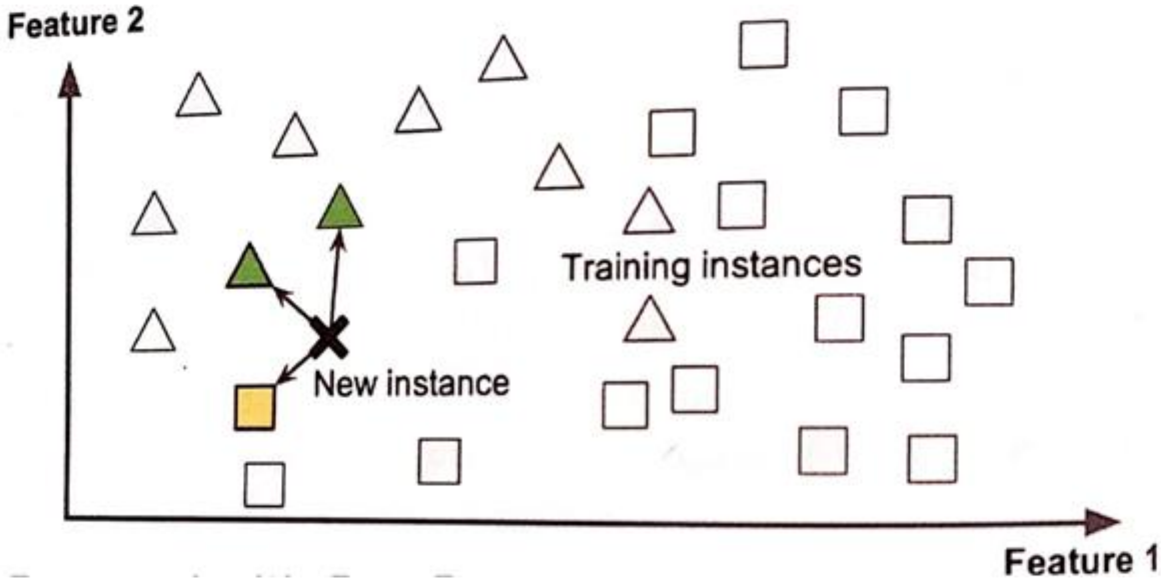
The machine learning NLP applications include, but are not limited to, classifying various texts into their respective categories, automatically flagging offensive comments on a public form or flagging spam email, summarizing long documents automatically and even creating human-like chatbots for specific purposes like customer support of a product. Each application utilizes a different machine learning algorithm such as Recurrent Neural Networks (RNNs).

The task for the project was to classify various texts pertaining to health-related Sustainable Development Goals (SDGs), outlined by the United Nations (UN). An example of the text is 'Quality Improvement Initiatives for Diabetes' and it's corresponding SDG would be '3.4.1 - Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease'. This is an example of a text that only has one label. This is not the case for the entire dataset.

Most applications of machine learning are geared towards mutually exclusive classification. Implying that it can only have one type of label at a time. The prototype programmed in the Python language for the project is a general solution for all multi-class and multi-label texts. Meaning that a single piece of text can have more than one label at a time and the possibility of labels are more than one two. For example, an essay could refer to either poverty or employment or both.

The task at hand had 27 possible labels with no upper limit to how many of the SDGs are incorporated in a particular piece of text. The dataset that we were working on had up to 10 labels simultaneously, implying that a single piece of text is related to 10 distinct SDGs simultaneously. The text was muddled with HTML code and other ASCII characters not containing useful information. To deal with this data mining problem the Natural Language ToolKit 3 (NLTK) library for python was used. NLTK, originally created in 2001, is a comprehensive natural language processing library that includes vital functions used in almost all text analytics projects. Popular functions such as removing stop words (punctuation or common words that do not add information to the sentence such as 'the', 'a', 'if', 'then' etc.), lowercasing the entire text and removing unknown words such as HTML code. These functions are used every time a natural text is analyzed to extract the valuable information from the clutter.

The model solves the problem in a generalized manner because the features used to train to model is the frequency of words categorized by label. This is an example of Instance Based Learning (IBL). IBL is a trivial form of learning which is to simply 'learn by heart' and then generalizes to new cases by using a similarity measure to compare them to the learned examples (Geron, 2019). This means the model consumes examples with labels, then when a new instance is fed without a label, the algorithm evaluates the closest possible label.



The figure above is an illustration of instance-based learning. The new instance (black cross) will be classified as a triangle because most of the similar instances belong to that class. Where the features would be the frequency of particular words. The words would be chosen with the frequency of their occurrence with respect to a particular label. For example, biomedical terms like ‘neuroscience’ were a pivotal feature for the label ‘3.b.2 - Total net official development assistance to medical research and basic health sector’. This is expected considering the SDG 3.b.2 pertains to texts discussing medical research. Once the features (popular words with respect to the label) are selected, when a new instance is processed, and it contains pivotal features, a classifier will be able to assign a probability of whether the text contains the label or not.

The classifier used to generate these probabilities in the developed model is the Naïve Bayes Classifier. In this multi-class classifier, every feature is incorporated into the model. This classifier first calculates the prior probabilities of the label. This is determined by the popular words. The more common the word, the higher the prior probability. When a new instance is fed, the words are examined and a likelihood estimate for each label is calculated. The label with the highest probability is assigned to the text.

The working of the Naïve Bayes algorithm implies that for the model developed, the different labels were created for a single Naïve Bayes Classifier. This is untrue because the classifier outputs a single label meaning the labels are mutually exclusive. To deal with the issue, a separate classifier is instantiated for every single label. This translates into every classifier deciding whether a label is triggered for a text or it is not, making each classifier a binary classifier. Finally, to make the classifier into a multi-class and multi-label classifier, an ensemble of all 27 label classifiers was pieced together. Finally, a single piece of text is processed by every single Naïve Bayes model outputting a true/false for every label.

2900 pieces of texts were used to train the models with each label getting its own set of texts. Individual Training accuracies of some of the models are shown below.

Label	Train Accuracy	Number of Texts
3.d.1	0.43	217
3.8.1	0.47	531
3.7.1	0.51	189
3.7.2	0.53	165
3.9.2	0.55	218
3.b.3	0.58	397
3.3.2	0.64	174
3.3.3	0.65	158
3.3.5	0.66	156
3.1.1	0.75	218
3.4.1	0.75	485
3.c.1	0.75	232
3.b.2	0.78	1044

Some labels train fine, others either underfit or overfit. The accuracy of these models can be vastly improved using better data wrangling techniques. Using more regular expressions to remove unknown words that the model can be sensitive towards.

The ensemble was then used to test 100 instances of data it had never seen before from the same dataset. The accuracy was a ratio calculated using the number of labels predicted correctly to total actual labels. This gave us an accuracy of ~62%. While the method to calculate the accuracy was more lenient than conventional means, it was used as a quick and general idea of how the model was performing. Since this is not the final face of the ensemble and more work is required till it is ready for launch.

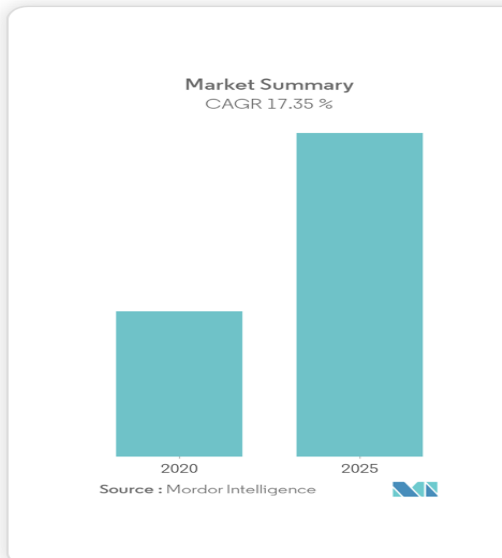
2. MARKET AND COMPETITION

2.1. Market size

The global text analytics market is expected to reach a value of USD 14.84 billion dollars by 2026 at a compound annual growth rate of 17.35 percent.

TEXT ANALYTICS MARKET - GROWTH, TRENDS, AND FORECAST (2021 - 2026)

Market Snapshot



Study Period:	2018 - 2026
Base Year:	2020
Fastest Growing Market:	Asia Pacific
Largest Market:	North America

Source: Mordor Intelligence

2.2. VeriText versus Competition

Text mining and classification software have been emerging through the years. Some of our main competitors are Keatext, InMoment, DiscoverText and MonkeyLearn, Text2Data, NetOwl, RapidMiner.

Although these text mining softwares have a similar purpose of automating business workflows with the help of machine learning capabilities, they differ in several ways as well. For instance, DiscoverText allows a customer to attach memos to documents and datasets. It removes duplicates and allows people to generate efficient summaries and detailed reports. MonkeyLearn, our closest competitor in concept, is another software that is partially customizable since it allows customers to classify text but with a limit of 500,000 queries per year and 10 custom models.

Our competitors mostly have specific target models and limited customization options. While studying our competitors and reading customer comments online came to our attention one of the complaints about the text mining software InMoment, where customers were discontent with not having some features they would like.

The following was a review from a user in retail:

“What do you dislike?”

“There are some things that would make sense that they have missed. For example, there is no feature that links the scores to the comments. So, we have to deduce which comment was lined up with which scores instead of knowing for sure”

Source: g2.com, InMoment Reviews

This is where our main advantage over our competitors comes in. Veritext is fully customizable to all customer needs; moreover our text classification algorithm is adaptable to different types of businesses and markets . Moreover we don’t limit our customers to a certain query number as they decide and shape the platform as they wish. This enables us on the contrary of our competitors to target all market segments equally.

Competitor Target Customers

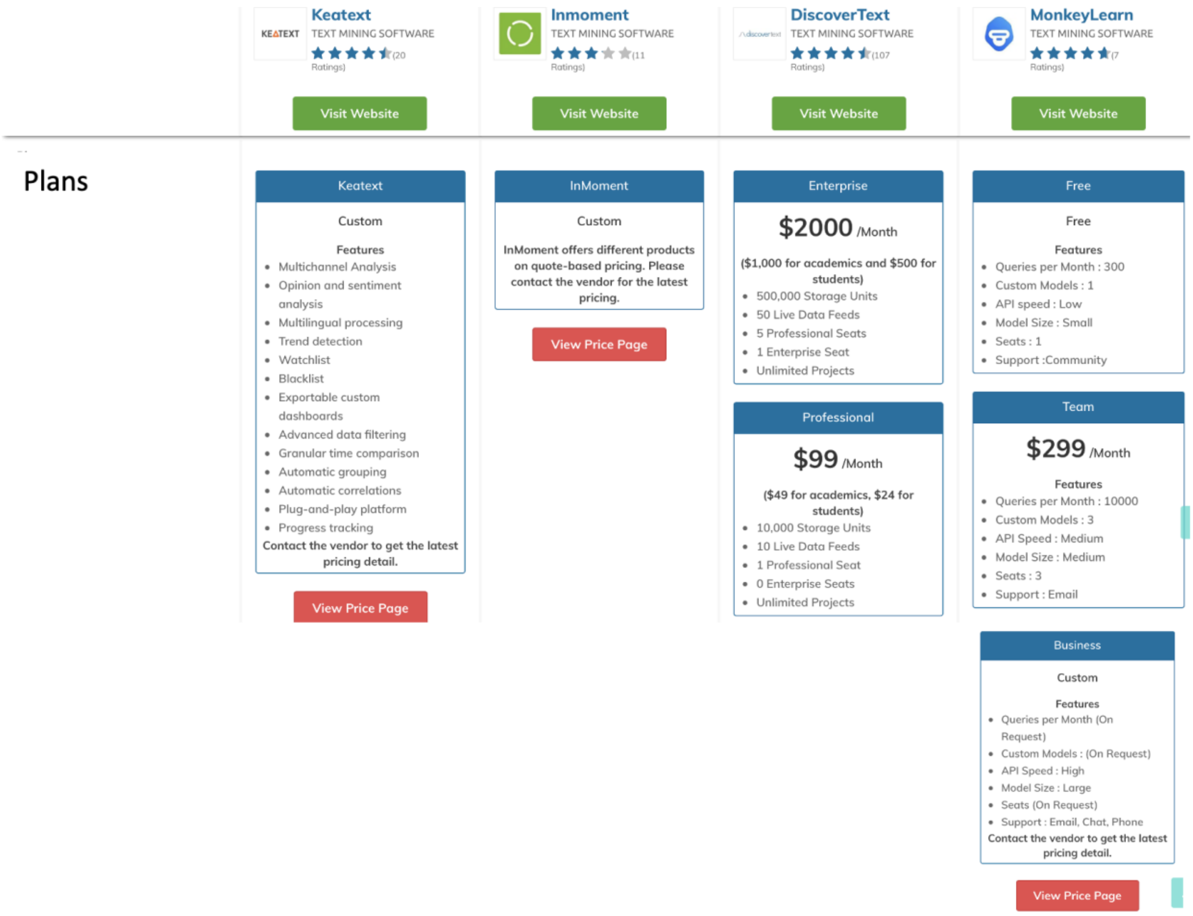
	Keatext TEXT MINING SOFTWARE (20 Ratings) Visit Website	Inmoment TEXT MINING SOFTWARE (11 Ratings) Visit Website	DiscoverText TEXT MINING SOFTWARE (107 Ratings) Visit Website	MonkeyLearn TEXT MINING SOFTWARE (7 Ratings) Visit Website
Customers				
Individuals	✗	✗	✗	✗
Freelancers	✗	✗	✗	✓
Large Enterprises	✗	✓	✗	✗
Medium Business	✓	✓	✓	✓
Small Business	✓	✗	✓	✓

Source: SaaSworthy

2.3. Pricing

We aim on implementing a subscription fee which is what some of our competitors use. Concerning pricing, our prices which are around $3500/month are almost similar to those of our competitor’s semi-customized platforms for businesses, but with way more added value. For instance, Keatext charges around $1000/month but only collects customer feedback in order to audit customer satisfaction as well as to identify emerging trends. We on the other hand provide full customization from types of models, number of models and even server size. Our software also has no seat limit like our competitors, which is how many people can use the software. This gives flexibility for the company to give access to all it’s team. These options we provide allow us to stand out among competitors.

Competitor Pricing Plans:



Source: SaaSworthy

Veritext Versus Competitor: Pricing and Customization

Product	Customization Level (out of 10)	Price per month($)
MonkeyLearn	5	6000
DiscoverText	3	2000
Keatext	0	1000
InMoment	2	6000
Veitext	10	3500
	X	Y

2.4. Future Target Markets

Because of the complete customization option Veritext would be able to fully cater for more market segments in the future such as insurance companies and legal practices. Our first target will be banks, our research leads us to see immense market potential and fast growth in this segment.

3. BUSINESS MODEL

3.1. Problem

During the course of their business, professionals have to deal with reading a lot of text to aid the quality of their work. For some, these texts can be voluminous and time consuming. This can affect the quality of work as emotions and human errors can affect the quality of output. This kind of errors of omission and commission has led many organisations to pay huge fines and penalties.

3.2. Solution

Over the last decade, a lot of attention is being given to RegTech. Regtech is the management of regulatory processes within the financial industry through technology. The main functions of regtech include regulatory monitoring, reporting, and compliance. However, VeriText is a level above just ensuring accuracy and monitoring through technology. With VeriText, a lot of text can be processed in a short time thereby saving time, cost and ensuring error and regulatory infractions are reduced significantly.

3.3. Unique Value Proposition

VeriText's unique value proposition is its distinctive customer experience. It will achieve this by customizing the model to the specific needs and requirements of each customer. VeriText will be the first to do this for its clients.

3.4. Unfair Advantage

VeriText currently does not have a wide network of possible stakeholders. Corporate relations still have to be developed. This could lead to a slower development compared to other companies who already have developed networks with established financial institutions.

3.5. Customer Segments

In the first few years, we are going to target large enterprises. The target market is the banking industry. However, the technology is adaptable in the future to different businesses and industries starting with legal and insurance companies.

3.6. Cost Structure

A major component of the cost structure for Veritext will be the cost of building and adapting the model to specific client needs. This will vary from client to client and will depend on the complexity of each client's operational demands.

3.7. Revenue Streams

Veritext's services will be required across several sectors of the economy. We will be rolling out with Banks and other financial institutions. We plan to extend Veritext's services to the Legal sector and Insurance sector within 5 years. We believe these sectors have a major need for Veritext's services and have the means to afford it. We carried out a market survey to find out the cost we will be saving potential clients with Veritext and discounted this amount to arrive at an average annually projected revenue per client of $40,000. Also, we will adopt an annual subscription revenue generation model which will require clients to renew their subscription annually to continue to access Veritext.

4. ROLLOUT PLAN

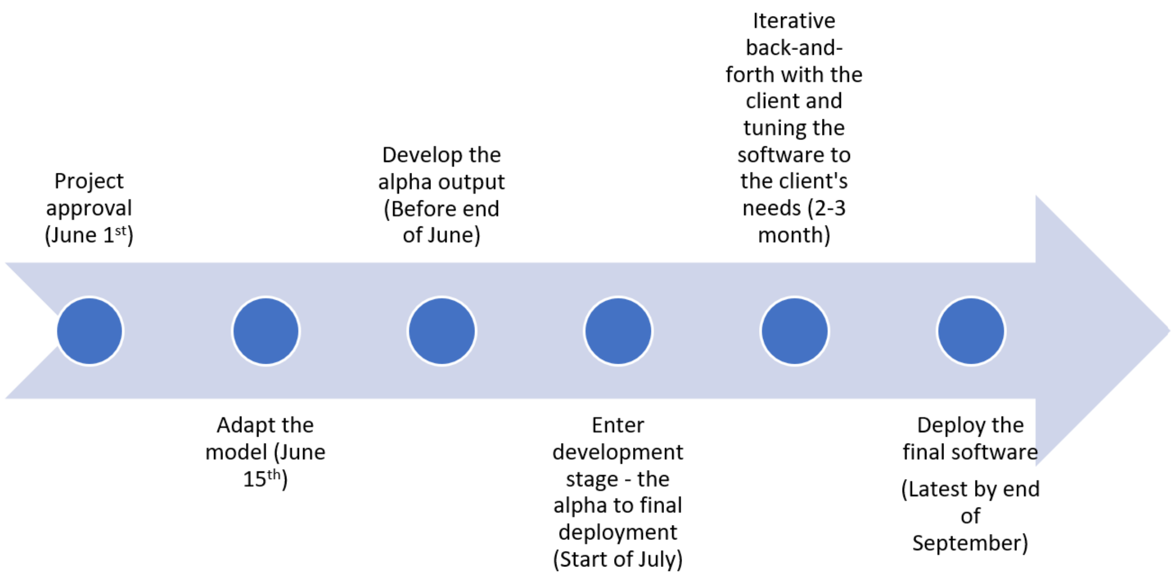
The software is aimed towards the legal teams of banks. Therefore, it is paramount to approach from the end user's perspective. Developing empathy for non-technical personnel and keeping the user experience (UX) of the final product intuitive and friendly. This attitude will allow the development team to get it right the first time. This is vital because even if the product has state of the art technology embedded into it, if the users do not use it; it is just an expensive paper weight.

Another feature of the rollout program would be that every single aspect of the software will be a liquid entity. Implying nothing will be hardcoded. This dynamic nature of the software will smooth the iterative nature of the development process and would be easy to implement clients' feedback.

The concrete plan is set up with the above outlined empathetic mindset. Two hypothetical timelines are shared below. First depicting the overall rollout of the entire company and a second timeline to show the individual deployment of the software.

Rollout as a Company

Rollout as an individual product



5. OPPORTUNITIES AND RISKS

The treatment of financial data is an important content that must be as accurate as possible because failure might lead to costly loss. Therefore, during the creation of personalized models for companies, we will be continually in contact with the companies to reach the best model accuracy possible thanks to training test data provided by companies. Depending on the sector and needs of companies, the algorithm's robustness may differ, and the margin of error will be discussed case by case.

In the best case, this approach will ensure strong models and create the best solutions for their needs and reduce possible risks due to lack of accuracy. Furthermore, such an approach to the problem might, in the worst case, not be sufficient. In the beginning, clients will prefer a less precise information classification approach with an over-representation of false positives to avoid losing information. Customer accumulation data from the same sectors will then improve and increase the robustness of the model to help the software be more accurate day after day and provide only data/information needed.

The need for privacy that companies with sensitive information/data required are tackled with private offline server setups to ensure access to the software. This possibility will provide risks for clients from cyber-attack our company may face.

6. FINANCIALS AND OTHER KEY METRICS

To generate the financial projections for Veritext, we have the following assumptions:

7.1 Customer Acquisition

New Customers	Q1	Q2	Q3	Q4
Year1	0	1	1	1
Year2	2	1	2	1
Year3	2	2	2	2
Total Customers	Q1	Q2	Q3	Q4
Year1	0	1	2	3
Year2	5	6	8	9
Year3	11	13	15	17

We are projecting to sign on 3 clients by the end of the first operating year and grow this number to 9 by the end of the second year and 17 by the end of the third year. This we will achieve by drawing on the combined depth of the network of the 5 founders of Veritext and the connections already made in the process of building this concept.

7.2 Financial Assumptions

Veritext's financial model is built on the following assumptions. We prepared two scenarios for our financial projections - Base Case Scenario and the Best-Case Scenario. The fundamental difference between these scenarios is the projected revenue per customer. Based on our market research and interviews, we are on the opinion that our model will be providing a solution that will be saving potential customers an average of $150,000 annually. Hence, we are projecting an average annual revenue of $25,000 per customer for the Base Case Scenario and $40,000 per customer for the Best-Case Scenario.

To build our financial model, we use the following financial metrics:

1. **Revenue**

Based on the projected average annual revenue and the frequency of customer acquisition, Veritext will have a revenue of $37,500 in its first year. This will grow to $175,000 in Year 2, $350,000 in Year 3, $525,000 in Year 4 and $787,500 in Year 5 for the Base Case Scenario. For the Best-Case Scenario, the company will have a revenue of $60,000 in its first year. This will grow to $280,000 in Year 2, $560,000 in Year 3, $840,000 in Year 4 and $1.26million in Year 5. As the global economy recovers from the economic downturn caused by COVID-19 pandemic, corporate revenues of companies will begin to rebound and companies can spend more. This will help our sales as our potential customers will have more money to spend.

2. **Cost of Goods Sold**

This is the direct cost associated with Veritext's model and it is the cost of adapting the model to each customer's specification. Using the current average industry rates, the cost of goods sold for Veritext will be $11,250 in Year 1, $21,040 in Year 2 and 3, while in Year 4 and 5, the cost of goods sold will grow to $31,560 and $47,340 respectively.

3. **Gross Margin**

The projected figures for Revenue and Cost of Goods Sold above will give us a Gross Margin of 69% in Year 1, 88% in Year 2 and 94% in Year 3, 4 and 5.

4. **Selling General and Administrative Expenses**

For the first 5 years of Veritext's operations, we have identified the following expenses:

Payroll - The company will start with the 5 founders as employees for the first year and will be on a lean salary of $30,000 per annum for the first 5 years. From the second year, we will be hiring a Software Developer and a Data Engineer with projected salaries of $60,000 per annum.

Rent - For the first year we plan to operate from the garage of one of the founders. We plan to get an office in the second year.

Web Services and Maintenance - This is the cost of hosting Veritext's platforms and after sale maintenance.

Utilities - Payment for Internet service, Phone bills, electricity and other utilities are recorded here.

Office Suppliers - These expenses include payment for office stationeries and other supplies to be used at the office.

7.3 Scenario Analysis

Based on the financial metrics above, kindly see a snapshot of our projected profit & loss statement for both the base and best cases:

<u>Base Case</u>

	Year 1	Year 2	Year 3	Year 4	Year 5
Revenues	$37,500	$175,000	$350,000	$525,000	$787,500
- COGS (Model Adaptation)	$11,520	$21,040	$21,040	$31,560	$47,340
Gross Margin	**$25,980**	**$153,960**	**$328,960**	**$493,440**	**$740,160**
SGA					
Payroll	$150,000	$270,000	$270,000	$270,000	$270,000
Payroll Tax	$7,500	$13,500	$13,500	$13,500	$13,500
Rent	$0	$30,000	$30,000	$30,000	$30,000
Web Services & Maintenance	$6,240	$6,240	$6,240	$6,240	$6,240
Utilities	$3,600	$3,600	$3,600	$3,600	$3,600
Office Suppliers	$3,600	$3,600	$3,600	$3,600	$3,600
Sub-total	**$170,940**	**$326,940**	**$326,940**	**$326,940**	**$326,940**
Profit/Loss Before Tax	**-$144,960**	**-$172,980**	**$2,020**	**$166,500**	**$413,220**
- Tax	-$11,597	-$13,838	$162	$13,320	$33,058
Profit/Loss After Tax	**-$133,363**	**-$159,142**	**$1,858**	**$153,180**	**$380,162**

Best Case

	Year 1	Year 2	Year 3	Year 4	Year 5
Revenues	$60,000	$280,000	$560,000	$840,000	$1,260,000
- COGS (Model Adaptation)	$11,520	$21,040	$21,040	$31,560	$47,340
Gross Margin	**$48,480**	**$258,960**	**$538,960**	**$808,440**	**$1,212,660**
SGA					
Payroll	$150,000	$270,000	$270,000	$270,000	$270,000
Payroll Tax	$7,500	$13,500	$13,500	$13,500	$13,500
Rent	$0	$30,000	$30,000	$30,000	$30,000
Web Services & Maintenance	$6,240	$6,240	$6,240	$6,240	$6,240
Utilities	$3,600	$3,600	$3,600	$3,600	$3,600
Office Suppliers	$3,600	$3,600	$3,600	$3,600	$3,600
Sub-total	**$170,940**	**$326,940**	**$326,940**	**$326,940**	**$326,940**
Profit/Loss Before Tax	**-$122,460**	**-$67,980**	**$212,020**	**$481,500**	**$885,720**
- Tax	-$9,797	-$5,438	$16,962	$38,520	$70,858
Profit/Loss After Tax	**-$112,663**	**-$62,542**	**$195,058**	**$442,980**	**$814,862**

7.4 Free Cash Flow Analysis

Based on these scenarios, Veritext will breakeven and be profitable from the 3rd year. A detailed break even analysis over a range of average revenue per customer and projected number of customers is shown below:

		Average yearly revenue per customer										
	($107,460)	$25,000	$26,000	$27,000	$28,000	$29,000	$30,000	$35,000	$40,000	$42,000	$43,500	$45,000
Average number of customers	2	($132,460)	($130,460)	($128,460)	($126,460)	($124,460)	($122,460)	($112,460)	($102,460)	($98,460)	($95,460)	($92,460)
	3	($107,460)	($104,460)	($101,460)	($98,460)	($95,460)	($92,460)	($77,460)	($62,460)	($56,460)	($51,960)	($47,460)
	4	($82,460)	($78,460)	($74,460)	($70,460)	($66,460)	($62,460)	($42,460)	($22,460)	($14,460)	($8,460)	($2,460)
	5	($57,460)	($52,460)	($47,460)	($42,460)	($37,460)	($32,460)	($7,460)	$17,540	$27,540	$35,040	$42,540
	6	($32,460)	($26,460)	($20,460)	($14,460)	($8,460)	($2,460)	$27,540	$57,540	$69,540	$78,540	$87,540
	7	($7,460)	($460)	$6,540	$13,540	$20,540	$27,540	$62,540	$97,540	$111,540	$122,040	$132,540
	8	$17,540	$25,540	$33,540	$41,540	$49,540	$57,540	$97,540	$137,540	$153,540	$165,540	$177,540
	9	$42,540	$51,540	$60,540	$69,540	$78,540	$87,540	$132,540	$177,540	$195,540	$209,040	$222,540
	10	$67,540	$77,540	$87,540	$97,540	$107,540	$117,540	$167,540	$217,540	$237,540	$252,540	$267,540
	11	$92,540	$103,540	$114,540	$125,540	$136,540	$147,540	$202,540	$257,540	$279,540	$296,040	$312,540
	12	$117,540	$129,540	$141,540	$153,540	$165,540	$177,540	$237,540	$297,540	$321,540	$339,540	$357,540
	13	$142,540	$155,540	$168,540	$181,540	$194,540	$207,540	$272,540	$337,540	$363,540	$383,040	$402,540
	14	$167,540	$181,540	$195,540	$209,540	$223,540	$237,540	$307,540	$377,540	$405,540	$426,540	$447,540
	15	$192,540	$207,540	$222,540	$237,540	$252,540	$267,540	$342,540	$417,540	$447,540	$470,040	$492,540
	16	$217,540	$233,540	$249,540	$265,540	$281,540	$297,540	$377,540	$457,540	$489,540	$513,540	$537,540
	17	$242,540	$259,540	$276,540	$293,540	$310,540	$327,540	$412,540	$497,540	$531,540	$557,040	$582,540
	18	$267,540	$285,540	$303,540	$321,540	$339,540	$357,540	$447,540	$537,540	$573,540	$600,540	$627,540
	19	$292,540	$311,540	$330,540	$349,540	$368,540	$387,540	$482,540	$577,540	$615,540	$644,040	$672,540
	20	$317,540	$337,540	$357,540	$377,540	$397,540	$417,540	$517,540	$617,540	$657,540	$687,540	$717,540

Based on the profit & loss statement above, a projected free cash flow statement was generated. Kindly see below:

	Year 1	Year 2	Year 3	Year 4	Year 5
Revenues	$37,500	$175,000	$350,000	$525,000	$787,500
- COGS (Model Adaptation)	$11,520	$21,040	$21,040	$31,560	$47,340
Gross Profit	$25,980	$153,960	$328,960	$493,440	$740,160
SGA	$170,940	$326,940	$326,940	$326,940	$326,940
EBITDA	($144,960)	($172,980)	$2,020	$166,500	$413,220
EBITDA Margin	-387%	-99%	1%	32%	52%
Depreciation	$0	$2,000	$2,000	$2,000	$2,000
Amortization	$0	$0	$0	$0	$0
EBIT	($144,960)	($174,980)	$20	$164,500	$411,220
Taxes (8%)	($11,597)	($13,838)	$162	$13,320	$33,058
Net Income	($133,363)	($161,142)	($142)	$151,180	$378,162
Depreciation & Amortization	$0	$2,000	$2,000	$2,000	$2,000
Capital Expenditure	$0	$0	$0	$0	$0
Additions to Intangibles	$0	$0	$0	$0	$0
Change in working capital	$0	($25,778)	$161,000	$151,322	$226,982
Free Cash Flow	($133,363)	($184,920)	$162,858	$304,502	$607,145

This statement shows that Veritext will start having a positive free cash flow from the 3rd year.

7.5 Enterprise Valuation

Using a Weighted Average Cost of Capital of 15%, Terminal Growth rate of 4% and Perpetuity Multiple rate of 11%, we have an enterprise valuation of $3.6million for Veritext as seen below.

Perpetuity Growth Method					
WACC	15%				
Terminal Growth Rate	4%				
Perpetuity Multiple	11.00				
Projected Years	1	2	3	4	5
FCF	($133,363)	($184,920)	$162,858	$304,502	$607,145
NPV	($115,968)	($139,826)	$107,082	$174,100	$301,858
NPV of 5-years FCF	**$327,246**				
Terminal Value of FCF	**$6,678,593**				
PV of TV (Discounted 5 Years)	**$3,320,441**				
Enterprise Value	**$3,647,687**				

In terms of financing, Veritext will require working capital of $300,000 to survive the first 2 years when it will be making losses. The founders will be contributing a total sum of $100,000 as startup capital. Hence, Veritext will require an equity injection of $200,000.

8. CONCLUSION & FUTURE PROSPECTS

State-of-the-art deep learning algorithms to classify the texts - Our current approach utilises the Naive Bayesian approach where each word has a weight. Instead if we use a recurrent neural network (RNN), not only will the accuracy of the model increase dramatically. Currently RNNs are considered to be the best text classifiers. Although RNNS are computationally intensive, once the model has been trained, the classification is swift.

Incorporating an ‘Online Learning’ feature so the algorithm can learn from new data in real time - Currently the model is an offline system. This means that the entire model needs to first train offline and only then can the final output be used to classify text. This is also called ‘batch learning’. The downside is that these systems are unable to adapt to changes and the entire system needs to train with the entire data before it can be used again. Online learning will eliminate this problem.

Using more advanced machine learning algorithms to *summarise long documents* - RNNS can also be used to summarise long pieces of text without loss of vital information. This feature will further distinguish the VeriText as an out of the box solution.

AI-driven chatbots to assist users with their queries - This is a more complex challenge. Chatbots can only solve extremely simple problems in their current technological form. They incorporate many Natural Language Processing (NLP) components and even include Natural Language *Understanding* (NLU) modules.

REFERENCES

Ascent (2019) https://www.ascentregtech.com/blog/infographic-regulatory-knowledge-automation-for-lawyers-explained/

Deloitte (2017) https://www2.deloitte.com/content/dam/Deloitte/us/Documents/regulatory/us-regulatory-future-of-regulatory-productivity-powered-by-regtech.pdf

Edward Loper, S. B. (2009). Natural Language Processing with Python

FinTech Global (2018) https://fintech.global/globalregtechsummit/why-banks-need-regtech-to-make-transaction-monitoring-more-effective-efficient/

Georgiou, M. (2019) https://www.imaginovation.net/blog/ai-in-banking-jp-morgan-case-study-benefits-to-businesses/

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow

Pwc (2019) https://www.pwc.in/consulting/financial-services/fintech/fintech-insights/regtech-a-new-disruption-in-the-financial-services-space.html

Reuters (2017) https://www.reuters.com/article/us-banks-fines/banks-paid-321-billion-in-fines-since-financial-crisis-bcg-idUSKBN1692Y2

Shriff, R. (2019) https://towardsdatascience.com/natural-language-processing-in-banking-current-uses-7fbbaee837de

SaasWorthy https://www.saasworthy.com/compare/keatext-vs-inmoment-vs-discover-text-vs-monkeylearn?pIds=312,316,319,2242

Mordor Intelligence https://www.mordorintelligence.com/industry-reports/text-analytics-market

G2, Reviews
https://www.g2.com/products/inmoment/reviews#survey-response-3622422