

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 29-04-2020		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 10-Sep-2018 - 9-Sep-2019	
4. TITLE AND SUBTITLE Final Report: I-Corps: CDnet: Community Detection in Large Networks			5a. CONTRACT NUMBER W911NF-18-1-0437		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 106011		
6. AUTHORS			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Nevada - Las Vegas 4505 Maryland Parkway Box 454019 Las Vegas, NV 89154 -4019			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 73772-SD-ICR.4		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Justin Zhan
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 000-000-0000

RPPR Final Report
as of 29-Apr-2020

Agency Code:

Proposal Number: 73772SDICR

Agreement Number: W911NF-18-1-0437

INVESTIGATOR(S):

Name: Justin Zhan
Email: jzhan@uark.edu
Phone Number: 0000000000
Principal: Y

Name: Leith Martin
Email: leith@leithmartin.com
Phone Number: 2053054698
Principal: N

Name: Sanket Chobe
Email: sanketchobe@gmail.com
Phone Number: 7252661351
Principal: N

Organization: **University of Nevada - Las Vegas**

Address: 4505 Maryland Parkway, Las Vegas, NV 891544019

Country: USA

DUNS Number: 098377336

EIN: 886000024

Report Date: 09-Dec-2019

Date Received: 29-Apr-2020

Final Report for Period Beginning 10-Sep-2018 and Ending 09-Sep-2019

Title: I-Corps: CDnet: Community Detection in Large Networks

Begin Performance Period: 10-Sep-2018

End Performance Period: 09-Sep-2019

Report Term: 0-Other

Submitted By: Justin Zhan

Email: jzhan@uark.edu

Phone: (000) 000-0000

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1

STEM Participants: 1

Major Goals: The goal is to explore the possibility of commercializing CDnet software product developed through the previous funded project. The product provides precise community detection in large network with superior performance. Through the funded projects, we studied the structure and dynamics of network communities and invented novel efficient methods for detection of network communities and building predictive models of behavior of groups of people. Although the actions of a particular individual may be too difficult to model, data mining and machine learning can be applied to large groups or ensembles, which can yield effective models and useful predictions. Our product CDnet has three main functions: (1) network community discovery and detection, 2) deep learning for network dynamics, and (3) a game-theoretic model of community evolution. This product has led to new insights into community structure of large networks, new theoretical models for understanding such structures and algorithms to harness them, and novel discoveries about connectivity structure of groups in networks. Such discoveries allow a company to explore its dynamics of network community structure and evolution.

Accomplishments: We have conducted more than 100 interviews for potential customers. The project followed along the curriculum guidelines set by I-Corps, where the team will focus on the lessons taught and then quickly engages with industry/customer. The first task was to immediately identify customers/industries that utilize networks and would find advantages of having a CDnet. This approach will require determining the customers' current limitations with existing technology, and whether switching the technology is justifiable. The second task was to determine the strategic market value of the technology. The goal is to determine that the proposed CDnet technology fit the customers' needs and provide sufficient benefits to warrant the use of the CDnet.

At the current stage, the proof-of-concept and prototype of the CDnet technology are well established by Chobe and Zhan. The current state of the technology is summarized as follows:

- Proof of concept with detailed synthesis and characterizations have been published.

RPPR Final Report as of 29-Apr-2020

- Three different CDnets have been tested with different properties.

At the end of the project, a combination of an edited video that illustrates the simplicity and use of CDnet for specific customer, an actual product that can be displayed to peers/customers, and/or pitch talk slide presentation for recruiting third party investors. A spin-off company with a website that highlights the technology, customer testimonies, and expanded product lines would also supplement the final demonstration if the final vote is to commercialize this CDnet technology. The final demonstration should relay a clear message that the technology has been optimized as commercial product, the product(s)/technology are well received by the customers, the projected revenue model entices the third party investors, and the expansion of the technology in other applications are apparent.

Training Opportunities: We have gone through the professional I-Corp training provided by experts.

Results Dissemination: We have attended a number of conferences to conduct interviews.

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI

Participant: Justin Zhan

Person Months Worked: 2.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Funding Support:

Participant Type: Faculty

Participant: Leith Martin

Person Months Worked: 2.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Funding Support:

Participant Type: Graduate Student (research assistant)

Participant: Sanket Chobe

Person Months Worked: 2.00

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Funding Support:

ARTICLES:

RPPR Final Report as of 29-Apr-2020

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Journal of Big Data
Publication Identifier Type: DOI **Publication Identifier:** <https://doi.org/10.1186/s40537-019-0243-y>
Volume: 6 **Issue:** 83 **First Page #:** 1
Date Submitted: **Date Published:** 4/27/19 9:23PM
Publication Location:
Article Title: Advancing community detection using Keyword Attribute Search
Authors: Sanket Chobe, Justin Zhan
Keywords: community detection, Keyword Attribute Search
Abstract: As social network structures evolve constantly, it is necessary to design an efficient mechanism to track the influential nodes and accurate communities in the networks. The attributed graph represents the information about properties of the nodes and relationships between different nodes, hence, this attribute information can be used for more accurate community detection. Current techniques of community detection do not consider the attribute or keyword information associated with the nodes in a graph. In this paper, we propose a novel algorithm of online community detection using a technique of keyword search over the attributed graph. First, the influential attributes are derived based on the probability of occurrence of each attribute type-value pair on all nodes and edges, respectively. Then, a compact Keyword Attribute Signature is created for each node based on the unique id of each influential attribute. The attributes on each node are classified into different classes, and thi
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: Y

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: PLoS ONE
Publication Identifier Type: DOI **Publication Identifier:** <https://doi.org/10.1371/journal.pone.0198066>
Volume: 13 **Issue:** 7 **First Page #:**
Date Submitted: **Date Published:** 4/29/18 4:31PM
Publication Location:
Article Title: Mining Association Rules for Low Frequency Itemsets
Authors: Jimmy Ming-Tai Wu , Justin Zhan , Sanket Chobe
Keywords: Mining, Association rules, Low-Frequency itemsets
Abstract: High utility itemset mining has become an important and critical operation in the Data Mining field. High utility itemset mining generates more profitable itemsets and the association among these itemsets, to make business decisions and strategies. Although, high utility is important, it is not the sole measure to decide efficient business strategies such as discount offers. It is very important to consider the pattern of itemsets based on the frequency as well as utility to predict more profitable itemsets. For example, in a supermarket or restaurant, beverages like champagne or wine might generate high utility (profit), but also sell less frequently compared to other beverages like soda or beer. In previous studies, it is observed that people who buy milk, bread, or diapers from a supermarket, also tend to buy beer or soda. But the items like milk, diapers, beer, or soda generate less utility (profit value) compared to beverages like champagne or wine. If we combine items like champa
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: Y

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: University of Nevada Las Vegas
Date Received: **Completion Date:** 4/30/18 4:15PM
Title: ADVANCING COMMUNITY DETECTION USING KEYWORD ATTRIBUTE SEARCH
Authors: Sanket Chobe
Acknowledged Federal Support: N

RPPR Final Report
as of 29-Apr-2020

Team CDNet





Winner Predictions
Betting Predictions
Attract Players

Our Initial Idea..

A software for social media marketing/analytics companies to find relevant customers and their features for precise marketing and fraud detection...

	Interviews	Count
	In Person	80
	Internet Call	14
	Phone Call	07
	Total	101



Team CDNet

Team 1492



Dr. Justin Zhan
Principal Investigator
PhD, Professor
Dept. of Computer Science



Leith Martin
Business Mentor
Executive Director,
Center for Entrepreneurship and Innovation

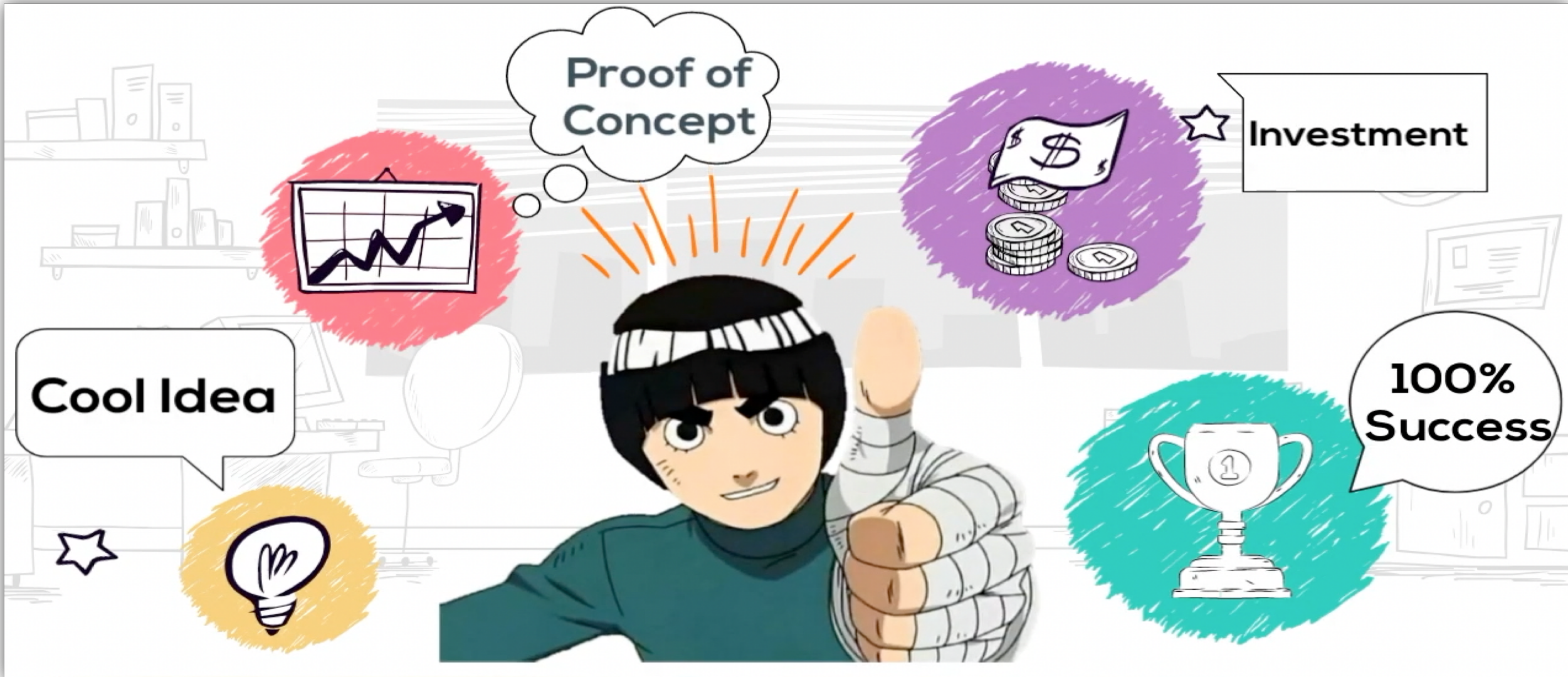


Sanket Chobe
Entrepreneur Lead
MS Graduate
Dept. of Computer Science



What we thought?

We have a cool idea..
We just need investment and right customers..



innovation within

Org: NSF I-Corps - Cohort: NSF I-Corps 2019 Winter Bay Area Cohort - Team: 1492_CDNet BMC

Key Partners	Key Activities	Value Propositions	Customer Relationships	Customer Segments
Infrastructure for cloud based solution.	Infrastructure for data ingestion and visualization.	1. Targeted marketing and advertising	Package or Plugin.	Network analysis companies
Open source technologies	Datasets for solution testing.	2. Advanced analytics	Optimized performance.	Ex: LinkedIn, Facebook, Instagram, Twitter
Visualiztion technologies	Domain expertise	3.Outlier or intrusion detection	Trust and information security	Who rely on network analysis technologies.
Domain specific knowledge providers	Optimize the performance	4. Risk analysis	Domain expertise	Amazon, Walmart, etc.
Potential investors for standalone product	Key Resources	5. Relevance and trust score for profiles	Channels	Customers who have not explored these areas and can benefit
	Datasets from various domains.		Solution to existing problem	Ex: esports analytics and gaming, Casino and hospitality industry
	Data storage platform with ingestion capabilities		Standalone solution to new problem	
	Data visualization platform		Demo to showcase solution	
	Domain experts, technical experts			
Cost Structure		Revenue Streams		
Expensive: resources for cloud based solutions		Integration/plugin with existing software		
Expensive: resources for Data visualization and infrastructure to provide support for real time visualization of the network		Companies targeting audience on stand alone platform		
Less expense: Data sets are available online, but getting data from customers might need some resources and expenses		Targeted audience or end users who will use our services directly		
Less expense: domain and technical expertise		Marketing for a Cloud and visualization based solution partners		
		Revenues from potential investors		

Value Propositions



- Precise marketing and targeted Advertisement



- Advanced Analytics



- Outlier and Intrusion detection
- Risk and Fraud Analysis
- Relevance & Trust Score for customer profiles

First BMC



I will be the greatest
Ninja ever...

Customer Segments



- Social network based companies



- Financial Companies for fraud detection



MGM RESORTS
INTERNATIONAL®



CAESARS
ENTERTAINMENT®

- Companies new to the social network and media analytics

INTERVIEWS..

Talked to more than
40 people ..

facebook

LinkedIn

NETFLIX

VISA

Morgan Stanley



The Struggle..

Team 1492

What is your product?
Who is your customer?

Whom should I talk to?
What questions should I ask them?



Linked in
NETFLIX

We have lot
of teams of
genius
people who
solve our
problems..



It's so difficult
to convince
them that we
have a great
idea !!

Updated BMC

Value Propositions



- 20% Improvement in accuracy of identifying new customers
- 20% Improvement in finding influential metrics for marketing



- 20-30% reduction in cost/time of integration and analysis of data
- 20% Increase in impact factor of the graph analytics tools

**Quantify Value Proposition..
Focus on Customer Roles..**



Customer Segments



- Product Engineers at Neo4j and Linkurious graph analytics tools.



MGM RESORTS
INTERNATIONAL®



CAESARS
ENTERTAINMENT®

- Director of Business Innovation & Strategies
- Director of Business Intelligence and Marketing
- Data Scientists and Data Analysts

Updated BMC

Team 1492

innovation within

Org: NSF I-Corps - Cohort: NSF I-Corps 2019 Winter Bay Area Cohort - Team: 1492_CDNet BMC

Key Partners	Key Activities	Value Propositions	Customer Relationships	Customer Segments
Cloud based infrastructure companies like Cloudera, AWS, Microsoft Azure.	Infrastructure for data ingestion and visualization.	20% improvement in accuracy of precise marketing	Package or Plugin.	Director of Business Development & Strategy planning at Casinos
Open source technologies	Datasets for solution testing.	20-30% improvement in accuracy of influential and correlated profiles	Optimized performance.	Director of Business Analytics and Data Science in Casinos
Visualization technologies	Domain expertise	20% reduction time and cost to identify the relevant data and parameters for the analytics.	Trust and information security	Data Analysts and Data Scientists in Casinos - Marketing, Business Intelligence department
Domain specific knowledge providers	Optimize the performance	20% improvement in the data quality for analytics	Domain expertise equipped with the advanced technological solution Channels	Marketing Manager in Casinos and Hotels
Potential investors for standalone product	Key Resources	20-30% increase in Outlier/fraud detection	Google Play Store/ Apple App Store	Data Analysts in esports and gaming
	Datasets from various domains.	80-90% accurate trust scores on profiles for risk analysis	Google, Facebook social media	Data Scientists in Facebook, LinkedIn, Twitter
	Data storage platform with ingestion capabilities	20% increase in the impact factor for the tool user	Problem solving for specific use-case through Demos	Product Engineers in Neo4j, Tiger Graph
	Data visualization platform		Conference, Press/ Networking events	Risk/ Fraud Detection Analysts in finance companies like Visa, Morgan Stanley
	Domain experts, technical experts			
Cost Structure		Revenue Streams		
Expensive: resources for cloud based solutions		Integration/plugin with existing software		
Expensive: resources for Data visualization and infrastructure to provide support for real time visualization of the network		Companies targeting audience on stand alone platform		
Less expense: Data sets are available online, but getting data from customers might need some resources and expenses		Targeted audience or end users who will use our services directly as a social network based product		
		Marketing for a Cloud and visualization based solution partners		
Less expense: domain and technical expertise		Revenues from potential investors		

SOME POSITIVE OUTCOMES...



Social network analysis provides tons of opportunities in predictive analytics..You guys are going in right direction..

Marc Smith: Director at NodeXL



Instead of finding what people are doing at Casinos, find why they are doing it.. You are trying to solve a very unique problem.. Kudos and Happy to help!!

Sam McMullen
Technologist, CEO @Fivegen
Corporatate Strategist @Geex.gg



Key Learning Moment..



What if we
narrow down
to specific
customer and
use-case?

“Quantum carburetor?”
Morty, you can’t just add a sci-fi
word to a car word and hope it
means something!

...Huh, looks like something’s wrong
with the microverse battery.



Can our software be useful
to Casinos in any way?



Aha moments..



Dude this is cool, eSports area lacks such tool and we can help you to get data and experiment.. We can work together to commercialize this..

Andrew Baca
Director Business Innovation and
Strategy





Aha moments..

Caesars would love to experiment with your technology to learn insights around customers participating in eSports events..

Drew Brunson – eSports Lead

MGM definitely has the challenge of improving the quality of data analytics.. eSports is a niche area, we are definitely interested to see your results..

Ganesh Variar
Executive Director Business
Intelligence



innovation within [™]						Org: NSF I-Corps - Cohort: NSF I-Corps 2019 Winter Bay Area Cohort - Team: 1492_CDNet BMC					
Key Partners		Key Activities		Value Propositions		Customer Relationships		Customer Segments			
Cloud infrastructure companies like AWS, Microsoft.		Infrastructure for data ingestion and visualization.		20% improvement in accuracy of precise marketing		Package or Plugin.		Director of Business Development & Strategy planning at Casinos			
Data Storage platforms like Neo4j Graph DB, TigerGraph.		Datasets for solution testing.		20-30% improvement in accuracy of influential and correlated profiles		Free license and application for initial phase		Director of Business Analytics and Data Science in Casinos			
Graphics and visualization companies		Plan and look for investment for \$1000,000 for 1 year plan		20% reduction time and cost to identify the relevant data and parameters for the analytics.		Trust and information security		Data Analysts and Data Scientists in Casinos - Marketing, Business Intelligence department			
Potential investors and early start ups		Form a team of technical and business experts in various Development MVP and experiment with Casinos		20% improvement in the data quality for analytics		Additional functionalities like people analytics, online casino analytics for Casinos		Marketing Manager in Casinos and Hotels			
Consulting services for software development.		Key Resources		20-30% increase in Outlier/fraud detection		Channels		Data Analysts in esports and gaming			
Small scale cloud infrastructure and service companies like Calligo Solutions.		Datasets from various domains.		80-90% accurate trust scores on profiles for risk analysis		Google Play Store/ Apple App Store		Data Scientists in Facebook, LinkedIn, Twitter			
eSports analytics companies		Data storage platform with ingestion capabilities		20% increase in the impact factor for the tool user		Google, Facebook social media Problem solving for specific use-case through Demos		Product Engineers in Neo4j, Tiger Graph			
		Data visualization platform						Risk/ Fraud Detection Analysts in finance companies like Visa, Morgan Stanley			
Cost Structure				Revenue Streams							
Expensive: resources for cloud based solutions				Integration/plugin with existing software							
Expensive: resources for Data visualization and infrastructure to provide support for real time visualization of the network				Targeted audience or end users who will use our services directly as a social network based product							
Less expense: Data sets are available online, but getting data from customers might need some resources and expenses				Marketing for a Cloud and visualization based solution partners							
Less expense: domain and technical expertise				Revenues from potential investors							
				License fee - approx. \$60000 per year							
				Cloud based subscription - \$100000 per year							

Final BMC

Value Propositions



20% Improvement in accuracy of identifying new customers



20-30% improvement in the quality of data analytics



10-20% reduction in cost/time of integration and analysis of data



Our motto is to be stronger than yesterday, if we have to we will be stronger than half a day ago, even a minute ago..

Customer Segments



Director of Business Innovation & Strategies



Director of Business Intelligence and Marketing



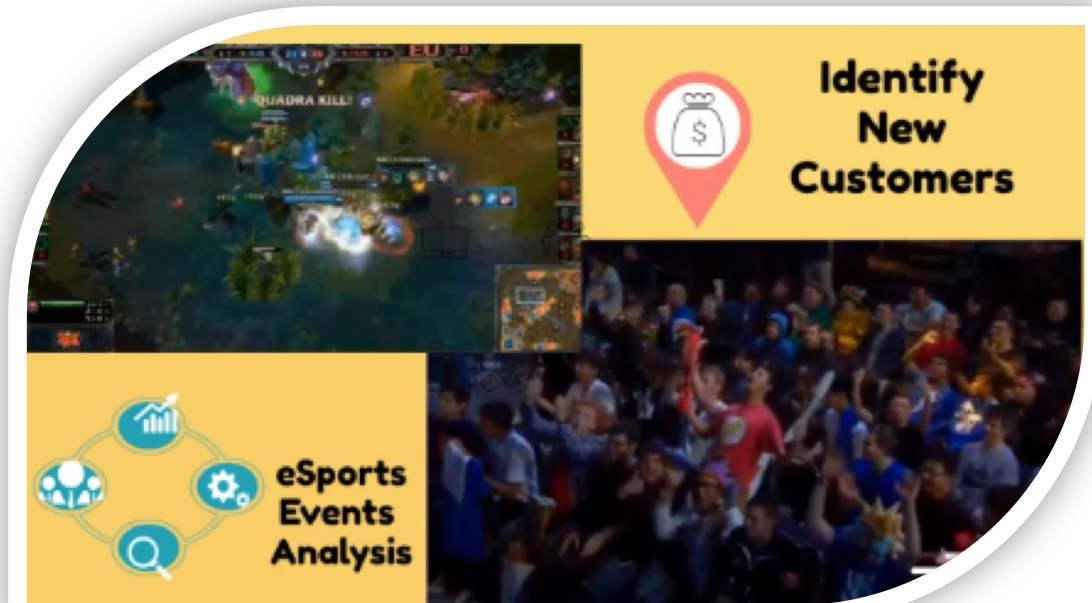
Data Scientists and Data Analysts



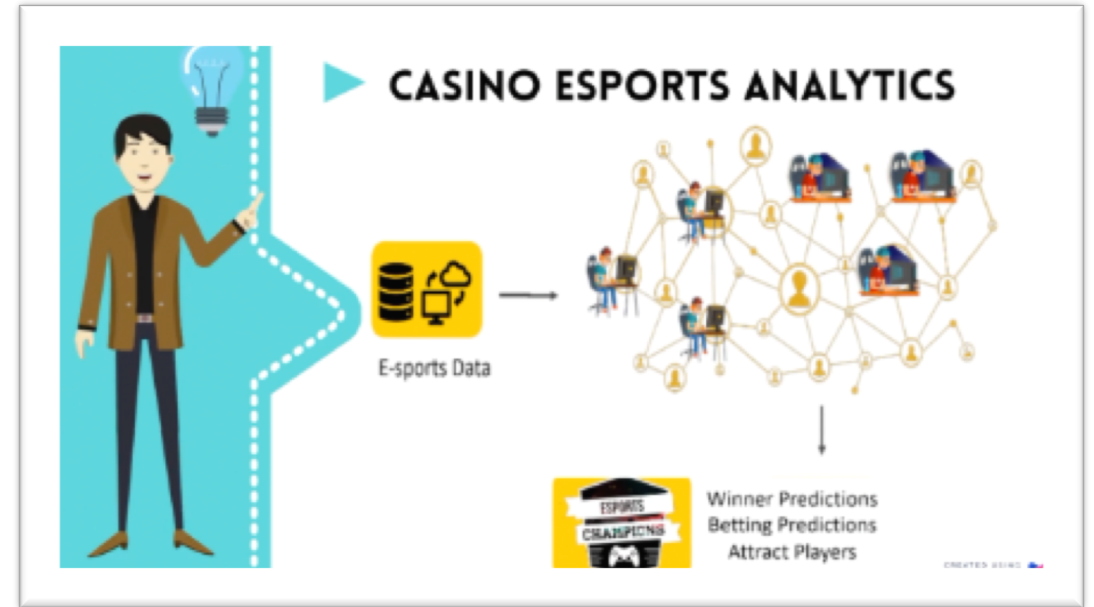
Data Analysts in eSports companies

This is Our Opportunity..

Our Customers



What they need...



We have the solution...

But...
It's a No GO...

- We see a product market fit..
- We may not have found an actual path but we see an exciting opportunity waiting for us..
- We are planning to figure out our path next ..



Talk to more
Casinos
2-3 months

Gather more
data around
Casinos and
their challenges

Proof of
Concept
6 months

Work with
Casinos on
experiments

Figure out the exact
Use-case Solution
1 year

Get comparison
results and
their evaluation

Thank You!!

If you believe in your dreams,
I will prove to you, that you can achieve
your dreams just by working hard.

Lessons Learned

<https://youtu.be/Wwg0piSqazA>

