Abstract: In this research work we developed a novel algorithm that can generate a narrative report of cricketing domain based on statistical data information extracted from raw values of match data. The algorithm is successful able to match quality standards of creating reporting with fair amount of increase in information gain.

Key Words: Natural language processing, Narratives science, Data, Information gain, Story.

I. INTRODUCTION
Narrative science is branch of NLP which turns data into stories. Narrative is defined as some kind of retelling, often in words (though it is possible to mime a story) of something that happened. Narrative meaning is about connections. It links individuals human actions and events into interrelated aspects of an understandable composite. Narrative science[1] is a kind of a tech solution that creates rich narrative content from data. Narrative are seamlessly created from structured data sources and can be fully customized to fit a customer’s voice, style and tone. Stories are created in multiple formats, including long form stories, headlines, tweets and industry reports with graphical visualizations. Multiple versions of the same story can be created to customize the content for each audience’s specific needs. In Narrative paradigm theory all meaningful communication is a form of storytelling or giving a report of events and so human beings experience and comprehend life as a series of ongoing narratives, each with their own conflicts, characters, beginnings, middles and ends. With help of narratives all forms of communication that appeal to our reason can best viewed as stories shaped by history culture and character and all forms of human communication are to be seen fundamentally as stories. The aim of narrative is to provide content and insight in those areas where it is either financially or logistically impossible for organizations to generate it themselves using traditional method.

Advantages of Narrative Science
Narrative provides the researcher with an understanding of data. Narrative gives the researcher access to stories or themes that the story teller may not even be conscious of. Narrative highlights changing perspectness and understanding of people and events as a function of time in the evaluation of an experience. Narrative science can help tackle a wide variety of business challenges and a broad range of company types e.g. Forbes.com uses narrative’s platform to generate corporate earnings preview stories. Narrative science helps companies leverage their data by automatically creating easy to use and consistent narrative reporting through our patented artificial intelligence platform. Narrative science turning big data into plain English and making sense of data. Big Data is defined as “data sets” whose size is beyond the ability of commonly used software tools to capture, manage and process the data within a tolerable elapsed time. Big Data is made of structured and unstructured information. 10% structured information is the data in databases and is about 10% of
story. Unstructured information is 90% of big data and is human information like emails, videos, tweets, Facebook posts, call-center conversations, closed circuit TV footage, mobile calls, website etc. Work with narrative science is easy to understand. With the help of narrative science we have more insights means deep knowledge in local language. Narrative science is more understanding and culturally customized.

Applications or why we use narrative science
We use narrative science to solve the problem of information explosion. People are essentially storytellers. Making decisions depends on judgements about ‘good reasons’. We are generating stories in the arenas of sports, finance, real estate and politics, when people experience a story, the phase of comprehension is where people form a mental representation about the text. The mental representation that is formed is called a situation model. While narrative science began its life providing content for media companies.

II. PREVIOUS WORK
In the previous work we work on a Narrator system for report generation in natural language. Narrator provides a set of report scenarios, based on numerical and textual indicators. It uses Google analytics data for a specific periods of time. The idea behind the developed system is based on merging statistical data with report templates predefined by the user. We can generate different kinds of reports with respect to various indicators, time periods and natural language processing.

III. METHODOLOGY OF WORK

- **Dataset of particular topic** - In this we can take dataset of sports data like cricket match. The column of the table represents matches and each row corresponds to attribute like date of match, day, place, teams, bowling teams, batting teams etc of given matches. Each value is known as a datum.
- **Extract statistical information** - In this step we can extract statistical information or we can say measure data in the arithmetic mean which is an average value for a group of numerical observation.
- **Develop formula based algorithm for the said event recorded in Dataset** - From extract data or information we can develop formula based algorithm. This algorithm is rule base algorithm in which our said event recorded in dataset.
- **Run story based on dataset** - In this step narration story is given out as output. Story is based on the dataset. Algorithm that extract the key facts and interesting insights from the data and transforms them into stories.
- **Calculate information gain** - Information gain is the expected reduction in entropy caused by partitioning the examples according to a given attributes. Entropy comes from information gain, higher the entropy the more the information content. Entropy is a common way to measure impurity. Entropy is very common in information theory characterizes the (im)purity of an arbitrary collection of examples.
IV. RESULTS

Figure 2 Information Gain

Entropy defines the purity of Information of an arbitrary collection of examples. Information Gain is the expected reduction in entropy earned by partitioning the examples according to a given attribute.

Figure 3 Information gain values

Narrative Coefficient = Relevant Narrative Score / Max. Narrative Score

Figure 4 Information gain before and after adding narration
Relavance Narration Score = Intro N + Body N + Conc N + Title N

![Figure 5 Information gain ratio](image)

Information Gain Ratio = Information gain/Intrinsic value = IG/Iv

V. CONCLUSIONS

All the results show that there is many folds of information gain when narration is added to dataset which are already based on the probability of events , most of the reports, journals, articles,today’s seems to have ‘formula’ or ‘plot’ in their expression ,these ‘formula’ or ‘plots’ can also be understood in terms of their mathematical existence, especially when data set is analysed on the basis of Descriptive statistics in which mean , average , frequency, mode min and max values can be translated into meanfull, along with its semantics narration which would be of machine quality but would be score fairly high in terms of qualitative perception of humans. By enhancing the information theory which is expressed in terms of probability now with narration of events captured in dataset ,we were able to generate automatic report tool which generates the report in HTML format.

REFERENCES