# Title: Learning from email content

## NER Feature 1: Extract calendar events from free-form emails

#### Team

|  |  |
| --- | --- |
| Name | Role |
| Adarsh Dattatri | Lead Data Scientist |
| Aditi Balaji | PM |
| Swami | Lead |
| Sudarshan | Engineering |

### Problem

Extract calendar events from free-form emails and autofill in Freshsales calendar

### Solution

Use facebook duckling to extract datetime text, detect meeting intent by using NLU models

### Key Results

Evaluated duckling output on ~50k contact emails across 4 accounts

##### Sample output:

|  |  |  |
| --- | --- | --- |
| Email received | Duckling output | Result |
| Could we move it to 2:30 at all? | NULL | False negative (FP) |
| Would you be able to do to the demo at 6:30 AM  or after 5 PM PST tomorrow? | [[{'datetime': '2019-03-23T06:30:00.000+00:00',  'time': '06:30:00',  'text': 'at 6:30 AM'},  {'datetime': '2019-03-23T01:00:00.000+00:00',  'time': '01:00:00',  'text': 'after 5 PM PST tomorrow'}]] | Mostly  true positives (TP) |
| I shared them via google doc yesterday, did you receive them? | [[{'datetime': '2019-03-21T00:00:00.000+00:00',  'time': '00:00:00',  'text': 'yesterday'}]] | False positive (FP) |
| That’s fantastic. Can you confirm that this price will be maintained after the first 12 months? | [[{'datetime': '2019-04-01T00:00:00.000+00:00',  'time': '00:00:00',  'text': 'after the first'}]] | False positive (FP) |

Duckling by itself doesn't do a great job at extracting calendar events. We are building rules and algorithms on top of duckling.

Sample output:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Email received | Duckling output | Result (duckling) | New algo output | Result (new algo) |
| Hi I spoke to Tanu yesterday to reschedule this call.  Can you do 1pm today UK time? | [[{'datetime': '2019-03-25T00:00:00.000+00:00',  'time': '00:00:00'  'text': 'yesterday'},    { 'datetime': '2019-03-26T13:00:00.000+00:00',  'time': '13:00:00',  'text': '1pm today UK time'}  ]] | FP and TP | { 'datetime': '2019-03-26T13:00:00.000+00:00',  'time': '13:00:00',  'text': '1pm today UK time'}  ]] | TP |

### Next Steps:

Train algorithms to extract calendar events from a subset of emails containing datetime text identified by duckling. Nextwealth to label these emails.

## NER Feature 2: Lead enrichment using email signature

### Team

|  |  |
| --- | --- |
| Name | Role |
| Adarsh Dattatri | Lead Data Scientist |
| Swami | Lead |
| Srivatsan | Product Lead |

### Problem:

Enriching lead information by NER on email signatures.

Solution:

TBD

## Identify parts of incoming email that are predictive of deal win

### Team

|  |  |
| --- | --- |
| Name | Role |
| Adarsh Dattatri | Lead Data Scientist |
| Rahul Kumar Sharma | Senior Data Scientist |
| Vishal Gupta | Intern |
| Swami | Lead |
| Srivatsan | Product Lead |

### Problem:

Identify parts of incoming email that are predictive of deal win as a feature and use it as:

1. Better insights for deal scores
2. Improve accuracy of deal scores
3. Detect email sentiment and display it alongside emails
4. Highlight important phrases/sentences in emails

### Samples:

|  |  |
| --- | --- |
| Email | Sentiment (towards Deal Closure) |
| We're **still interested in demoing freshsales** to the whole team. What times are you available this week? | Positive |
| We are just experiencing some difficulties on our end with getting the packing right and need another few weeks. If you can **give us another month for free** and the discount **ill sign up today**. | Positive |
| [alejandro@alohassandals.com](mailto:alejandro@alohassandals.com) **has accepted this invitation**.  Freshsales Discussion with Alejandro Hi there, Deivanai Annamalai is inviting you to a scheduled Zoom meeting | Positive |
| You can confirm 11 agents for now and **we will go with the Estate plan**. Jithin' | Positive |
| Sorry for not responding earlier to calls and email. We are in a very busy period and **haven't had a chance to focus on this system** to any great degree. Once we have a little breathing space I will be in touch regarding your solution | Negative |
| Apologies but neither of us have had time to look at this further yet. This is a **busy period for our business**. Rather than take up too much of your time, maybe we should look to **re-evaluate this in January** when schedules should be a little less busy | Negative |
| I still **didnt get much chance to set up yet**. I also compare with Salesforce, what is the benefit or **difference between your product and Salesforce**? | Negative |
| I am going to have to run some numbers, but at first glance this still **seems rather expensive** when I think about how much we plan to grow over the new few years. | Negative |

### Solution:

Currently, training supervised algorithms on parts of email/complete emails with paid wins as labels. The AUC from the basic model on this data is 63%.

### Next Steps:

1. The current approach labels the data as per deal closure outcome (paid-wins/loss). NextWealth to tag emails as per their sentiment.
2. Improve the model using new tagged data.
3. Use a sequence model to capture the context and negation of the emails.

#### 