

Google Cloud Platform Machine Learning Service  
- Prediction API를 중심으로 -

김진형  
2017.03.28.

## 1. Google

<https://www.google.com/>

## 2. Google product

<https://www.google.com/intl/en/about/products/>

## 3. Google product for Developers

<https://developers.google.com/products/?hl=kr>

## 4. Google product for Developers → Google Cloud Platform

<https://cloud.google.com/>

## 5. Google product for Developers → Google Cloud Platform → Machine Learning

<https://cloud.google.com/docs/>

## 6. Google product for Developers → Google Cloud Platform → Machine Learning → Prediction API

<https://cloud.google.com/prediction/docs/>

- Documentation
  - Home
  - Free Trial Guide
  - The Basics
  - Compare Platforms
  - Start by Programming Language
    - Tutorials and Solutions
  - Compute
  - Containers
  - Storage
  - Networking
  - Big Data
  - Machine Learning
  - Operations
  - Identity and Security
  - Developer Tools
  - Resource Management
  - Sample Code and Videos
  - Support and Billing
    - Enterprise Guide
    - Launch Checklist

## Learn about products and services

Cloud Platform offers a full spectrum of cloud products and services for compute, storage, networking, big data, machine learning, operations, and more.

**Compute**

- Choosing a Computing Option
- App Engine
- Compute Engine
- More

**Storage**

- Choosing a Storage Option
- Cloud Storage
- Cloud Spanner <sup>Beta</sup>
- More

**Networking**

- Cloud DNS
- Cloud Interconnect
- CDN Interconnect
- More

**Big Data**

- BigQuery
- Cloud Dataflow
- Cloud Dataproc
- More

**Machine Learning**

- Cloud Machine Learning <sup>Beta</sup>
- Vision API
- Speech API <sup>Beta</sup>
- More

**Operations**

- Stackdriver Logging
- Stackdriver Monitoring
- Stackdriver Error Reporting
- More

**Identity and Security**

- Auth Guide
- Cloud IAM
- Cloud KMS
- More

**Developer Tools**

- Cloud SDK
- Cloud APIs
- Cloud Shell
- Cloud Source Repositories
- More

**Resource Management**

- Deployment Manager
- Cloud Launcher
- Container Builder <sup>Beta</sup>
- More

More -> Prediction API

## API를 어떻게 이해할 것인가?

- 간단한 설명
- Prediction API를 기반으로
- Prediction API 사용법
- Prediction API로 이해하기

## API를 어떻게 활용할 것인가?

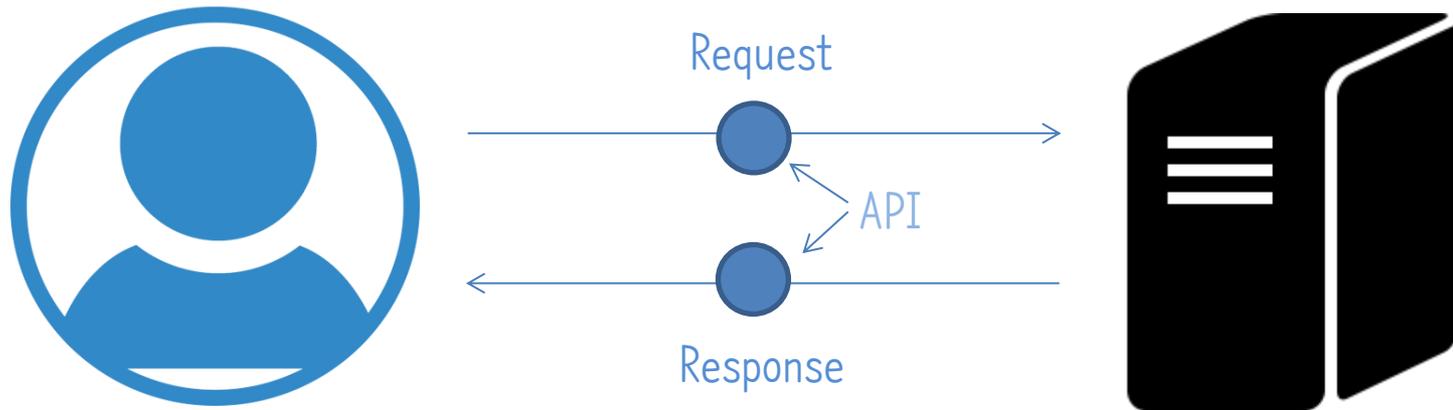
- Structuring Data
- Upload Data
- Authentication
- Node.js를 활용한 코드

## Google Cloud Platform Machine Learning API는?

API를 어떻게 이해할 것인가?

- 간단한 설명
- Prediction API를 기반으로
- Prediction API 사용법
- Prediction API로 이해하기

## Application Programming Interface(API)



# API를 어떻게 이해할 것인가? - Prediction API를 기반으로 (1/1)

- Cloud Prediction API
  - Product Overview
  - Documentation
  - Quickstart
- How-to Guides
  - All How-to Guides
  - Developer's Guide
  - Smart Autofill Spreadsheets Add-on
  - Use Cases
- APIs & Reference
  - All APIs & Reference
  - V1.6 (Latest)
  - Samples and Libraries
  - Prediction Gallery
  - PMML Support
  - Performance Tips
  - Batch HTTP Requests
- Resources
  - All Resources
  - Pricing
  - Glossary
  - Release Notes
  - Support
  - Service Level Agreement

Prediction API

☆☆☆☆☆  
의견 보내기

## Google Cloud Prediction API Documentation

Google Cloud Prediction API provides a RESTful API to build Machine Learning models. Prediction's cloud-based machine learning tools can help analyze your data to add various features to your applications, such as customer sentiment analysis, spam detection, recommendation systems, and more.

- Quickstart**  
Learn in 5 minutes
- How-to Guides  
Perform specific tasks
- APIs & Reference  
API references

- 이 페이지가 ☆☆☆☆☆ 의견 보내기
- 목차
  - Before you begin
  - Upload training data
  - Train the model
  - Confirm completion of training
  - Send a query
  - Clean up
  - What's next

- ### Prediction API 사용법
1. 가입, 결제정보 등록, 프로젝트 생성, Enable API
  2. Training data 업로드 (Storage API)
  3. Training model 생성
  4. Training model 완료 확인
  5. Prediction



### 목차

**Before you begin**

Upload training data

Train the model

Confirm completion of training

Send a query

Clean up

What's next

## Before you begin

1. Select or create a Cloud Platform project.

[GO TO THE PROJECTS PAGE](#)

2. Enable billing for your project.

[ENABLE BILLING](#)

3. [Enable the Prediction and Google Cloud Storage APIs](#) for your project. Because you will use the APIs Explorer for this quickstart, you don't need to go to the Credentials page afterward.
4. [Download the sample training data file](#) ( `language_id.txt` ). This file contains several text snippets and the language of each snippet (English, Spanish, or French) as CSV data.

<https://console.cloud.google.com/>

The screenshot shows the Google Cloud Platform console interface. At the top, there is a blue header with the Google Cloud Platform logo, a search bar, and several utility icons. Below the header, the main content area is divided into a left sidebar and a main workspace. The sidebar contains navigation links for IAM and Project management. The main workspace shows the 'Project' management page with a '+ 프로젝트 만들기' button highlighted in a red box. A modal dialog box titled '새 프로젝트' (New Project) is open, also outlined in red. It contains a text input field for the project name, a note stating '프로젝트 ID는 프로젝트 이름을 기반으로 합니다.' (Project ID is based on the project name.), and two buttons at the bottom: '취소' (Cancel) and '만들기' (Create).

Google Cloud Platform Project

IAM 및 관리자 프로젝트 + 프로젝트 만들기 프로젝트 삭제 정보 패널 숨기기

모든 프로젝트

IAM 할당량 서비스 계정 라벨 GCP 개인정보 보호 및 보안 설정 암호화 키

이름

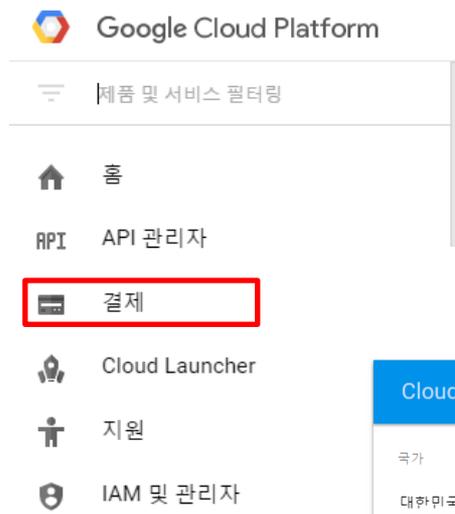
### 새 프로젝트

프로젝트 이름 ?

프로젝트 ID는 프로젝트 이름을 기반으로 합니다. ? 수정

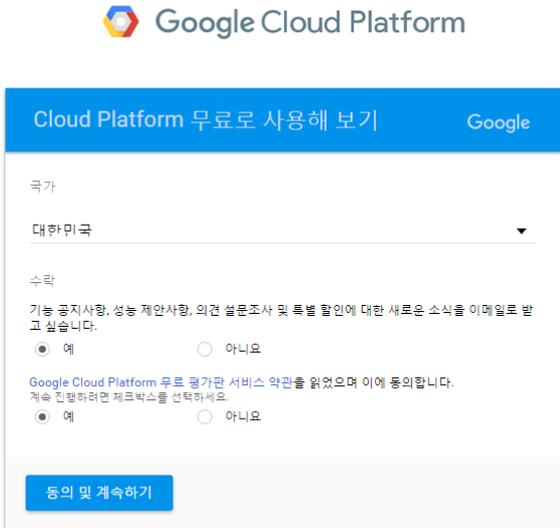
취소 만들기

# API를 어떻게 이해할 것인가? - Prediction API 사용법 1 (3/5)



Google Cloud Platform

- 제품 및 서비스 필터링
- 홈
- API 관리자
- 결제**
- Cloud Launcher
- 지원
- IAM 및 관리자



Google Cloud Platform

Cloud Platform 무료로 사용해 보기

국가: 대한민국

수락:  예  아니요

Google Cloud Platform 무료 평가판 서비스 약관을 읽었으며 이에 동의합니다. 계속 진행하려면 체크박스를 선택하세요.

예  아니요

[동의 및 계속하기](#)

[Privacy policy](#)



- 모든 Cloud Platform 제품에 액세스**  
앱, 웹사이트, 서비스를 구축하고 실행하는 데 필요한 Firebase, Google Maps API 등의 모든 기능을 이용할 수 있습니다.
- 무료 크레딧 300달러**  
가입하여 Google Cloud Platform에서 이후 12개월 동안 사용할 수 있는 300달러의 혜택을 받아 보세요.
- 무료 평가판 종료 후 자동 청구되지 않음**  
신용카드를 요청하는 이유는 자동 가입을 방지하기 위해서입니다. 무료 평가판 이용 중이나 종료 후에 요금이 청구되지 않습니다.



Google Cloud Platform

Cloud Platform 무료로 사용해 보기

고객 정보

계정 유형: 사업자

이름 및 주소: 트/시

주소 입력란 1

주소 입력란 2

연체명

이메일

유선 번호

기본 연락처

결제 옵션

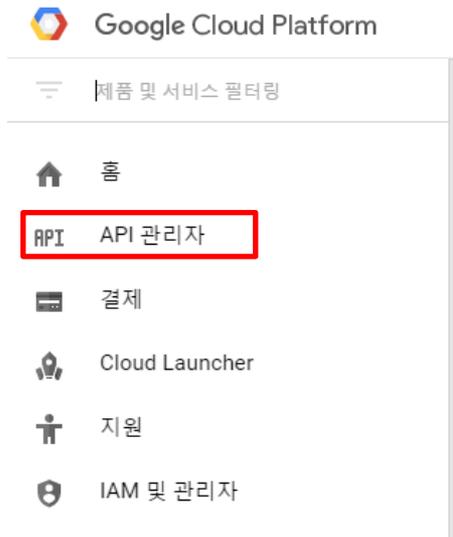
자세히 보기

결제 수단: 신용카드 또는 직불카드 주소가 위의 주소와 동일합니다.

[무료 평가판 시작하기](#)

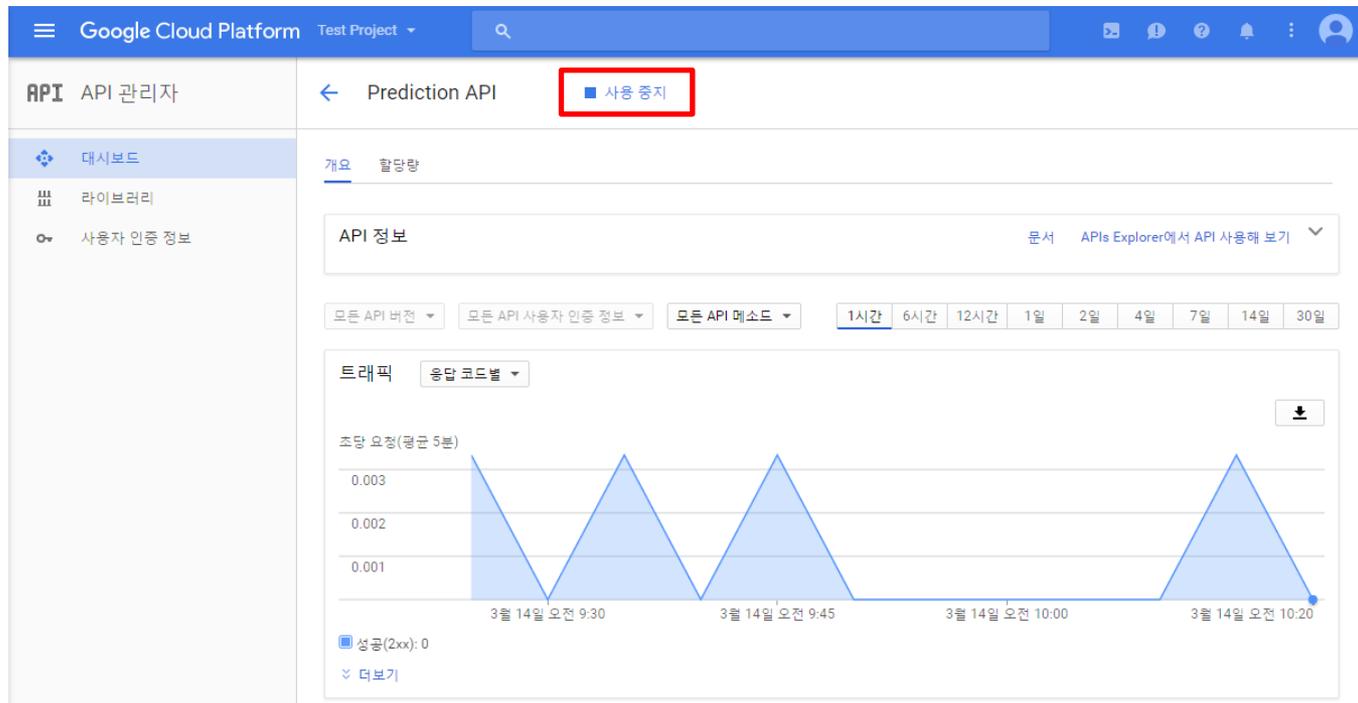
[Privacy policy](#)

# API를 어떻게 이해할 것인가? - Prediction API 사용법 1 (4/5)



Google Cloud Platform

- 제품 및 서비스 필터링
- 홈
- API API 관리자**
- 결제
- Cloud Launcher
- 지원
- IAM 및 관리자



Google Cloud Platform Test Project

API 관리자 Prediction API **■ 사용 중지**

개요 할당량

API 정보 [문서](#) [APIs Explorer에서 API 사용해 보기](#)

모든 API 버전 | 모든 API 사용자 인증 정보 | 모든 API 메소드 | 1시간 | 6시간 | 12시간 | 1일 | 2일 | 4일 | 7일 | 14일 | 30일

트래픽 응답 코드별

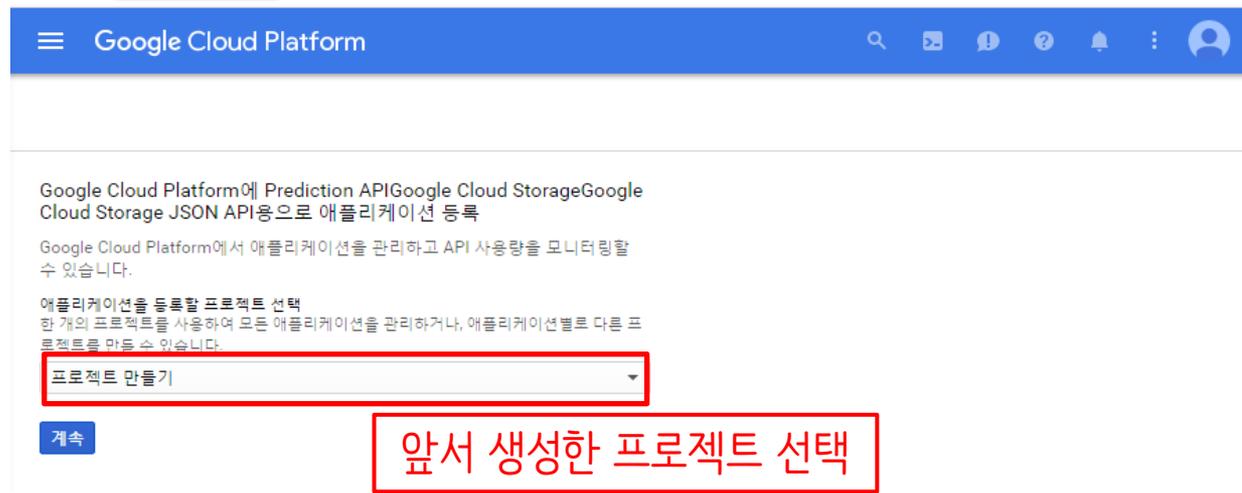
초당 요청(평균 5분)

| 시간              | 요청량   |
|-----------------|-------|
| 3월 14일 오전 9:30  | 0.003 |
| 3월 14일 오전 9:45  | 0.003 |
| 3월 14일 오전 10:00 | 0.000 |
| 3월 14일 오전 10:20 | 0.003 |

■ 성공(2xx): 0

▽ 더보기

닫기 [무료 평가판 신청](#)



Google Cloud Platform에 Prediction API, Google Cloud Storage, Google Cloud Storage JSON API용으로 애플리케이션 등록

Google Cloud Platform에서 애플리케이션을 관리하고 API 사용량을 모니터링할 수 있습니다.

애플리케이션을 등록할 프로젝트 선택  
한 개의 프로젝트를 사용하여 모든 애플리케이션을 관리하거나, 애플리케이션별로 다른 프로젝트를 만들 수 있습니다.

**프로젝트 만들기**

[계속](#)

**앞서 생성한 프로젝트 선택**

☰ Google Cloud Platform

Google Cloud Platform에 Prediction API, Google Cloud Storage, Google Cloud Storage JSON API용으로 애플리케이션 등록

Google Cloud Platform에서 애플리케이션을 관리하고 API 사용량을 모니터링할 수 있습니다.

애플리케이션을 등록할 프로젝트 선택  
한 개의 프로젝트를 사용하여 모든 애플리케이션을 관리하거나, 애플리케이션별로 다른 프로젝트를 만들 수 있습니다.

프로젝트 만들기

계속

뒤에서 좀 더 자세히..

## API가 사용 설정됨

Prediction API, Google Cloud Storage, Google Cloud Storage JSON API(가) 사용 설정되었습니다.

다음으로 API를 사용하려면 올바른 사용자 인증 정보가 필요합니다.

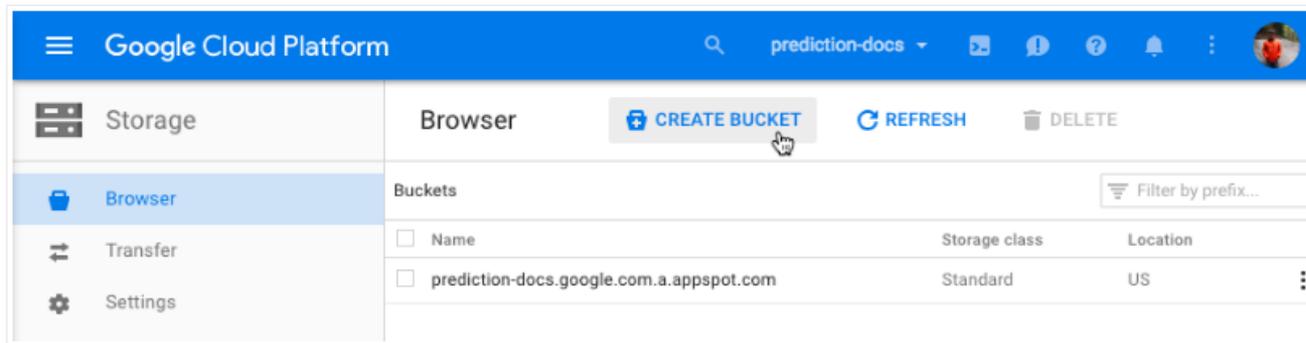
사용자 인증 정보로 이동

## Upload training data

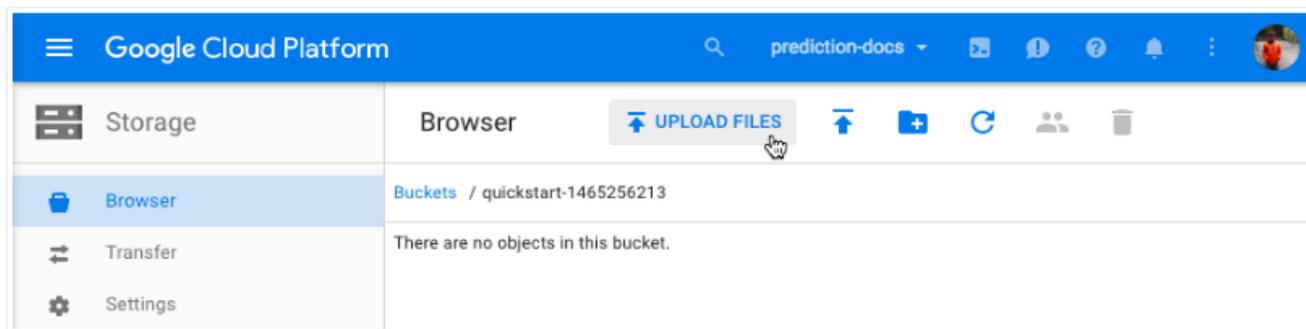
1. In the Google Cloud Platform Console:

[OPEN THE CLOUD STORAGE BROWSER](#)

2. Click **Create bucket** to create a new bucket.



3. Enter a globally unique name in the **Name** field. The following name has been generated for you if you'd like to use it: `quickstart-1488954768`. Or, pick your own.
4. Keeping the default values for the other fields, click **Create**.
5. After the bucket is created, click **Upload Files** and upload `language_id.txt`.



## Quickstart

Learn in 5 minutes

### 목차

Before you begin

**Upload training data**

Train the model

Confirm completion of training

Send a query

Clean up

What's next

## Train the model

To train the model, call the `prediction.trainedmodels.insert` method, passing a unique name for this predictive model, and the location of the training data.

```
POST https://www.googleapis.com/prediction/v1.6/projects/[PROJECT_ID]/trainedmodels
{
  "id": "language-identifier",
  "storageDataLocation": "quickstart-1465256213/language_id.txt"
}
```

Use the button below to send this request using the APIs Explorer. You must replace the following values in the Explorer:

- **project:** Your Cloud Platform Console [project ID](#).
- **Request body:** Update the value of `storageDataLocation` with your bucket name.

TRY IT!

A successful response looks like:

Click

```
{
  "kind": "prediction#training",
  "id": "language-identifier",
  "selfLink": "https://www.googleapis.com/prediction/v1.6/projects/prediction-docs/trainedmodels/language-identifier",
  "storageDataLocation": "quickstart-1465256213/language_id.txt"
}
```



## Quickstart

Learn in 5 minutes

### 목차

Before you begin

Upload training data

**Train the model**

Confirm completion of training

Send a query

Clean up

What's next

# API를 어떻게 이해할 것인가? - Prediction API 사용법 3 (2/2)

Services > Prediction API v1.6 > prediction.trainedmodels.insert

**project**

{your-project-id}

The project associated with the model. (string)

앞서 만든 프로젝트 ID 입력

**fields**

Selector specifying which fields to include in a partial response.

[Use fields editor](#)

**Request body**

```
{  
  "id": "language-identifier",  
  "storageDataLocation": "{your-bucket-name}/language-id.txt"  
}
```

만들고자 하는 모델의 ID

앞서 만든 버킷(스토리지) ID 입력

**bold red** = required

Authorize and execute

[Execute without OAuth](#)

Click -> Click

## Confirm completion of training

Use the `prediction.trainedmodels.get` method to check the status of training, passing the ID of the predictive model.

```
GET https://www.googleapis.com/prediction/v1.6/projects/[PROJECT_ID]/trainedmodels/language-identifie
```

Use the button below to send this request using the APIs Explorer. You must replace the following values in the Explorer:

- **project:** Your Cloud Platform Console [project ID](#).

TRY IT!

In the response, examine the `trainingStatus` property to see if the status is `RUNNING` or `DONE`:

```
{
  "kind": "prediction#training",
  "id": "language-identifier",
  "selfLink": "https://www.googleapis.com/prediction/v1.6/projects/prediction-docs/trainedmodels/langu
  "created": "2016-06-07T22:51:13.702Z",
  "trainingComplete": "2016-06-07T22:51:32.468Z",
  "modelInfo": {
    "numberInstances": "406",
    "modelType": "classification",
    "numberLabels": "3",
    "classificationAccuracy": "0.99"
  },
  "trainingStatus": "DONE"
}
```



## Quickstart

Learn in 5 minutes

목차

Before you begin

Upload training data

Train the model

Confirm completion of training

Send a query

Clean up

What's next

# API를 어떻게 이해할 것인가? - Prediction API 사용법 4 (2/2)

Services > Prediction API v1.6 > prediction.trainedmodels.get

|                |  |   |
|----------------|--|---|
| <b>project</b> | <input type="text" value="{your-project-id}"/>   | The project associated with the model. (string)   |
| <b>id</b>      | <input type="text" value="language-identifier"/> | The unique name for the predictive model. (string)  |
| <b>fields</b>  | <input type="text"/>                             | Selector specifying which fields to include in a partial response.<br><a href="#">Use fields editor</a> |

**bold red** = required

Execute

Click

앞서 만든 프로젝트 ID 입력

트레이닝하여 만든 모델의 ID 입력

## Send a query

After the training is complete, you can send queries to the service to be evaluated against the predictive model. To do so, call the `prediction.trainedmodels.predict` method, passing the name of the model and the query.

```
POST https://www.googleapis.com/prediction/v1.6/projects/prediction-docs/trainedmodels/language-ident

{
  "input": {
    "csvInstance": [
      "Sont des mots qui vont tres bien ensemble"
    ]
  }
}
```

Use the button below to send this request using the APIs Explorer. You must replace the following values in the Explorer:

- **project:** Your Cloud Platform Console [project ID](#).

TRY IT!

In the response, examine the `outputLabel` property to see what language the Google Prediction API thinks the string is in:

Click

```
{
  "kind": "prediction#output",
  "id": "language-identifier",
  "selfLink": "https://www.googleapis.com/prediction/v1.6/projects/prediction-docs/trainedmodels/langu
  "outputLabel": "French",
  "outputMulti": [
    {
      "label": "English",
      "score": "0.000000"
    },
    {
      "label": "French",
      "score": "1.000000"
    },
    {
      "label": "Spanish",
      "score": "0.000000"
    }
  ]
}
```

You can try sending queries in either of the other languages that the model was trained on (English and Spanish).



## Quickstart

Learn in 5 minutes

### 목차

Before you begin

Upload training data

Train the model

Confirm completion of training

Send a query

Clean up

What's next

# API를 어떻게 이해할 것인가? - Prediction API 사용법 5 (2/2)

Services > Prediction API v1.6 > prediction.trainedmodels.predict

**project**

{your-project-id}

The project associated with the model. (string)

앞서 만든 프로젝트 ID 입력

**id**

language-identifier

The unique name for the predictive model. (string)

트레이닝하여 만든 모델의 ID 입력

**fields**

Selector specifying which fields to include in a partial response.  
[Use fields editor](#)

Request body

```
{
  "input":
  {
    "csvInstance":
    [
      "Sont des mots qui vont tres"
    ]
  }
}
```

“예측”하고자 하는 문구 입력(영어, 스페인어, 불어 중에서)

**bold red** = required

Execute

Click

<https://developers.google.com/apis-explorer/>



prediction

Loading...



APIs Explorer



Services

Learn more about using the Prediction API by reading the [documentation](#).

All Versions

[Search Result](#) > Prediction API v1.6

Authorize requests using OAuth 2.0:  OFF

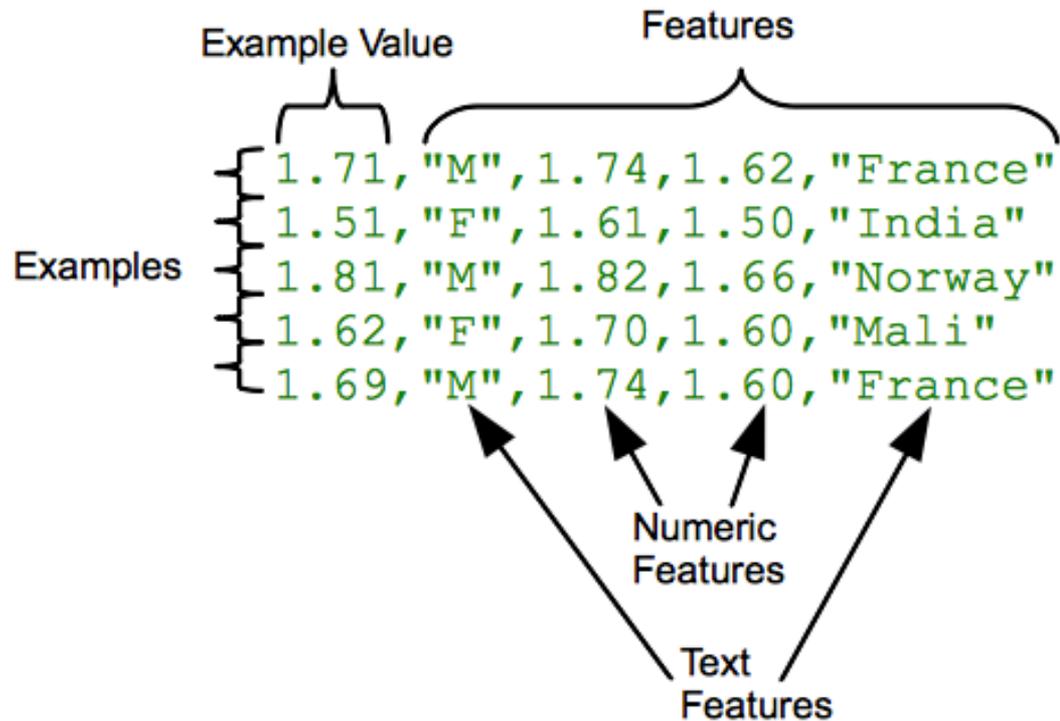
Request History

|                                  |  |
|----------------------------------|--|
| prediction.hostedmodels.predict  | Submit input and request an output against a hosted model.       |
| prediction.trainedmodels.analyze | Get analysis of the model and the data the model was trained on. |
| prediction.trainedmodels.delete  | Delete a trained model.  |
| prediction.trainedmodels.get     | Check training status of your model.                             |
| prediction.trainedmodels.insert  | Train a Prediction API model.                                    |
| prediction.trainedmodels.list    | List available models.   |
| prediction.trainedmodels.predict | Submit model id and request a prediction.                        |
| prediction.trainedmodels.update  | Add new data to a trained model.                                 |

API를 어떻게 활용할 것인가?

- Structuring Data
- Upload Data
- Authentication
- Node.js를 활용한 코드

## Sample Training Data



### Training Data File Format

Training data file is uploaded to Google Cloud Storage as a **CSV (comma-separated value) file**. Think of this file as a table, with each row representing one example, and commas separating columns. The first column is the example value, and all additional columns are features.

Empty cells will not cause an error, but you should avoid having empty cells because an empty string cell evaluates to "" (text features) or zero (numeric features), which will throw off the matching algorithm. There is no way to differentiate "unknown value" from zero or "" in the data.

After training a model from a data file, you can add additional training data to a model data using [streaming training](#). You can also delete model.

The CSV training file must follow these conventions:

- **Maximum file size is 2.5GB**
- You must have a minimum of **six examples** in your training file
- **No header row is allowed**
- Only one example is allowed per line. A single example cannot contain newlines, and cannot span multiple lines.
- Columns are separated by commas. Commas inside a quoted string are not column delimiters.
- **The first column represents the value (numeric or string) for that example. If the first column is numeric, this model is a regression model; if the first column is a string, it is a categorization model. Each column must describe the same kind of information for that example.**
- The column order of features in the table does not weight the results; the first feature is not weighted any more than the last.
- As a best practice, remove punctuation (other than apostrophes ' ) from your data. This is because commas, periods, and other punctuation rarely add meaning to the training data, but are treated as meaningful elements by the learning engine. For example, "end." is not matched to "end".

- **Text strings:**

- Place double quotes around all text strings.
- Text matching is case-sensitive: "wine" is different from "Wine."
- If a string contains a double quote, the double quote must be escaped with another double quote, for example: "sentence with a ""double"" quote inside"
- Strings are split at whitespace into component words as part of the prediction matching. That is, "Godzilla vs Mothra" will be split into "Godzilla", "vs", and "Mothra" when searching for closest matches. If you require a string to remain unsplit, such as for proper names or titles, use underscores or some other character instead of whitespace between words. For example: `Godzilla_vs_Mothra`. If you want to assign a set of labels to a single feature, simply include the labels: for example, a genre feature of a movie entry might be "comedy animation black\_and\_white". The order of words in a string also does not affect matching value, only the quantity of matched words.
- Quoting a substring does not guarantee an exact match. That is, placing quotes around the phrase "John Henry" will not force matches only with "John Henry"; "John Henry" has partial matches to both "John" and "Henry". However, more matches per string generates a higher score, so "John Henry" will match "John Henry" best.

- **Numeric values:**

- Both integer and decimal values are supported.
- **Numbers in quotes without whitespace will be treated as numbers, even if they are in quotation marks.** Multiple numeric values within quotation marks in the same field will be treated as a string. For example:
  - Numbers: "2", "12", "236"
  - Strings: "2 12", "a 23"

For categorization model

|    | A | B        | C        | D        | E        |
|----|---|----------|----------|----------|----------|
| 1  | F | 0.378487 | -0.01278 | 0.359555 | -1.79957 |
| 2  | F | -0.85324 | -1.51883 | 0.045366 | 0.001923 |
| 3  | T | 0.747537 | 0.868592 | -0.66966 | -1.70514 |
| 4  | T | 0.933162 | 0.079565 | -1.18398 | 0.8003   |
| 5  | F | -0.93291 | -0.37751 | 1.678622 | -2.11681 |
| 6  | T | 0.139463 | 0.496548 | -0.2819  | -0.09795 |
| 7  | F | -0.10949 | -0.54634 | 0.543921 | -0.18542 |
| 8  | T | 0.045765 | 0.562324 | -0.30935 | 0.777913 |
| 9  | F | 0.055782 | -0.7851  | 0.209391 | -0.1538  |
| 10 | T | 1.16569  | 0.690024 | 0.271404 | 0.216281 |
| 11 | T | -0.25918 | -0.07874 | -0.2608  | 0.516445 |
| 12 | T | 0.744942 | 1.346154 | -0.19878 | 0.76774  |

For regression model

|    | A        | B        | C        | D        | E        |
|----|----------|----------|----------|----------|----------|
| 1  | -2.04246 | 0.378487 | -0.01278 | 0.359555 | -1.79957 |
| 2  | -2.06037 | -0.85324 | -1.51883 | 0.045366 | 0.001923 |
| 3  | 0.145093 | 0.747537 | 0.868592 | -0.66966 | -1.70514 |
| 4  | 5.920947 | 0.933162 | 0.079565 | -1.18398 | 0.8003   |
| 5  | -7.5433  | -0.93291 | -0.37751 | 1.678622 | -2.11681 |
| 6  | 3.50033  | 0.139463 | 0.496548 | -0.2819  | -0.09795 |
| 7  | -1.09369 | -0.10949 | -0.54634 | 0.543921 | -0.18542 |
| 8  | 5.35553  | 0.045765 | 0.562324 | -0.30935 | 0.777913 |
| 9  | -0.26916 | 0.055782 | -0.7851  | 0.209391 | -0.1538  |
| 10 | 4.293475 | 1.16569  | 0.690024 | 0.271404 | 0.216281 |
| 11 | 3.984568 | -0.25918 | -0.07874 | -0.2608  | 0.516445 |
| 12 | 8.857899 | 0.744942 | 1.346154 | -0.19878 | 0.76774  |

# API를 어떻게 활용할 것인가? - Upload Data (1/1)

버킷 / quickstart-1488936542

| <input type="checkbox"/> 이름  | 크기       | 유형                       | 저장소 클래스        | 최종 수정 시간           |
|--|----------|--------------------------|----------------|--------------------|
| <input type="checkbox"/>  D_cat.csv       | 50.12MB  | application/vnd.ms-excel | Multi-Regional | 17. 3. 13. 오후 1:41 |
| <input type="checkbox"/>  D_lin.csv       | 59.84MB  | application/vnd.ms-excel | Multi-Regional | 17. 3. 13. 오후 1:41 |
| <input type="checkbox"/>  language_id.txt | 145.07KB | text/plain               | Multi-Regional | 17. 3. 8. 오전 10:54 |

2개 중 0개 업로드 완료

취소

D\_cat.csv

 25%

×

D\_lin.csv

 20%

×

<https://cloud.google.com/vision/docs/common/auth>  
[https://cloud.google.com/vision/docs/common/auth#set\\_up\\_a\\_service\\_account](https://cloud.google.com/vision/docs/common/auth#set_up_a_service_account)

## Set up a service account

The screenshot shows the Google Cloud Platform console interface. On the left is a navigation sidebar with 'API 관리자' selected. The main content area is titled '사용자 인증 정보' (User Authentication). A dropdown menu is open, showing options: 'API 키' (API Key), 'OAuth 클라이언트 ID' (OAuth Client ID), and '서비스 계정 키' (Service Account Key). The '서비스 계정 키' option is highlighted with a red box. Below the menu, there is a table of existing service account keys.

| 제한사항 | 키                                     |
|------|---------------------------------------|
| 없음   | AlzaSyCU7K0k30QrRJKp8uDppquOsiqmjOi6A |

| 서비스 계정 키  | 생성일         | 서비스 계정 |
|---|-------------|--------|
| <input type="checkbox"/> ID                                       |             |        |
| <input type="checkbox"/> 211cf2b39de8a56c6e9e78292a65dd856cb5f7f8 | 2017. 3. 8. | DEV    |



## 서비스 계정 키 만들기

서비스 계정

선택...

키 유형

비공개 키가 포함된 파일을 다운로드합니다. 이 키가 손실될 경우 복구할 수 없으므로 안전하게 저장하세요.

JSON

추천

P12

P12 형식을 사용하는 코드와의 하위 호환용

생성

취소

=> 복잡한 이름의 JSON 파일이 다운로드 됨

## <https://developers.google.com/identity/protocols/application-default-credentials>

### Authenticating with Application Default Credentials

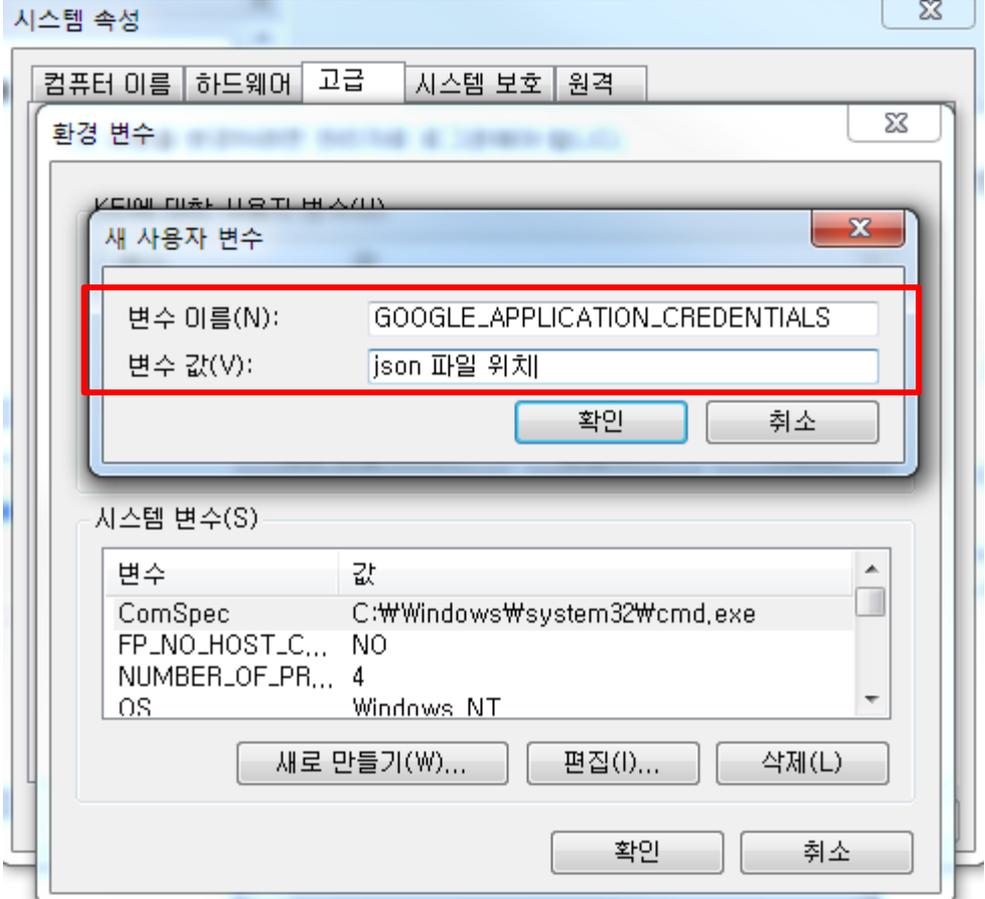
#### How the Application Default Credentials work

You can get Application Default Credentials by making a single client library call. The credentials returned are determined by the environment the code is running in. Conditions are checked in the following order:

1. The environment variable `GOOGLE_APPLICATION_CREDENTIALS` is checked. If this variable is specified it should point to a file that defines the credentials. The simplest way to get a credential for this purpose is to create a Service account key in the Google API Console:
  - a. Go to the [API Console Credentials page](#).
  - b. From the project drop-down, select your project.
  - c. On the Credentials page, select the **Create credentials** drop-down, then select **Service account key**.
  - d. From the **Service account** drop-down, select an existing service account or create a new one.
  - e. For **Key type**, select the **JSON** key option, then select **Create**. The file automatically downloads to your computer.
  - f. Put the \*.json file you just downloaded in a directory of your choosing. This directory must be private (you can't let anyone get access to this), but accessible to your web server code.
  - g. Set the environment variable `GOOGLE_APPLICATION_CREDENTIALS` to the path of the JSON file downloaded.
2. If you have installed the Google Cloud SDK on your machine and have run the command `gcloud auth application-default login`, your identity can be used as a proxy to test code calling APIs from that machine.
3. If you are running in Google App Engine production, the built-in service account associated with the application will be used.
4. If you are running in Google Compute Engine production, the built-in service account associated with the virtual machine instance will be used.
5. If none of these conditions is true, an error will occur.

앞서 다운로드 받은 \*.json 파일을 원하는 위치에 두고  
GOOGLE\_APPLICATION\_CREDENTIALS라는 이름의 환경변수를 생성

# API를 어떻게 활용할 것인가? - Authentication (4/4)



<https://github.com/keibigdata/GCPStudy>

```
29 function train (callback) {
30   auth(function (err, authClient) {
31     if (err) {
32       return callback(err);
33     }
34   }
35
36   var trainedmodels = google.prediction({
37     version: 'v1.6',
38     auth: authClient
39   }).trainedmodels;
40
41   trainedmodels.insert({
42     project: 'test-project-160901',
43     resource: {
44       "id": "random-regression",
45       "storageDataLocation": "quickstart-1488936542/D_lin.csv"
46     }
47   }, function (err, res){
48     if(err) console.log(err);
49     console.log(res);
50   });
51
52 });
53 }
```

```
D:##01 Study##07 GCP>node training.js
{ kind: 'prediction#training',
  id: 'random-regression',
  selfLink: 'https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-regression',
  storageDataLocation: 'quickstart-1488936542/D_lin.csv' }
```

## API를 어떻게 활용할 것인가? - Node.js를 활용한 코드 (2/5)

```
28 function get (callback) {
29   auth(function (err, authClient) {
30
31     if (err) {
32       return callback(err);
33     }
34
35     var trainedmodels = google.prediction({
36       version: 'v1.6',
37       auth: authClient
38     }).trainedmodels;
39
40     trainedmodels.get({
41       project: 'test-project-160901',
42       id: "random-regression"
43     }, function (err, res){
44       if(err) console.log(err);
45       console.log(res);
46     });
47
48   });
49 }
```

```
D:\W01 Study\W07 GCP>node get.js
< kind: 'prediction#training',
  id: 'random-categorization',
  selfLink: 'https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-categorization',
  created: '2017-03-14T23:51:54.731Z',
  trainingComplete: '2017-03-14T23:53:11.351Z',
  modelInfo:
    < numberInstances: '7000000',
      modelType: 'classification',
      numberLabels: '3',
      classificationAccuracy: '0.94' >,
  trainingStatus: 'DONE' >
```

```
D:\W01 Study\W07 GCP>node get.js
< kind: 'prediction#training',
  id: 'random-regression',
  selfLink: 'https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-regression',
  created: '2017-03-13T05:46:51.717Z',
  trainingStatus: 'RUNNING' >
```

```
D:\W01 Study\W07 GCP>node get.js
< kind: 'prediction#training',
  id: 'random-regression',
  selfLink: 'https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-regression',
  created: '2017-03-13T05:46:51.717Z',
  trainingComplete: '2017-03-13T07:44:29.128Z',
  modelInfo:
    < numberInstances: '10000000',
      modelType: 'regression',
      meanSquaredError: '1.01' >,
  trainingStatus: 'DONE' >
```

```
28 function predict (callback) {
29   auth(function (err, authClient) {
30
31     if (err) {
32       return callback(err);
33     }
34
35     var trainedmodels = google.prediction({
36       version: 'v1.6',
37       auth: authClient
38     }).trainedmodels;
39
40     trainedmodels.predict({
41       project: 'test-project-160901',
42       id: "random-regression",
43       resource: {
44         input: {
45           csvInstance: [-2.1, 3.1, 4.1, 3.1]
46         }
47       }
48     }, function (err, res){
49       if(err) console.log(err);
50       console.log(res);
51     });
52
53   });
54 }
```

```
77 var parser = parse({delimiter: ','}, function (err, data) {
78   async.eachSeries(data, function (line, callback) {
79     predict(function(){
80       callback();
81     }, line);
82   }, function(err){
83     var file = fs.createWriteStream('array.csv');
84     file.on('error', function(err) { console.log(err); });
85
86     resultArr.forEach(function(v) {
87       console.log(v);
88       file.write(v.join(', ') + '\n'); });
89     file.end();
90   })
91 })
92
93 var stream = fs.createReadStream(inputFile).pipe(parser);
94
```

시행착오...

## Linear Data

- 연속형 데이터 추정일 경우 100만 건을 학습시켰을 때 시간이 오래 걸림(정확히는 알 수 없으나 3시간 정도 걸린 것 같음..)
- 70만 건을 학습시켰을 때, time limit exceeded 에러가 나면서 학습이 되지 않음
- 새로운 데이터에 대한 적용 -> 포기(죄송  $\pi\pi$ )

```
D:\W01 Study\W07 Node.js>node get.js
< kind: 'prediction#training',
  id: 'random-regression',
  selfLink: 'https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-regression',
  created: '2017-03-14T06:43:27.656Z',
  modelInfo: < numberInstances: '0' >,
  trainingStatus: 'ERROR: TRAINING TIME LIMIT EXCEEDED' >
```

## Binary Data

- 70만 건을 학습시켰을 때, 다행히 시간이 오래 걸리지 않음(10분 내외)
- 정확도도 0.94로 아주 높음
- 그러나 나머지 데이터에 적용했을 때는..?
- 5000건에 대해서 수행했을 때, 4699건을 정확하게 예측하여 정확도 0.9398로 거의 정확하게 예측

Google Cloud Platform Machine Learning API는?

## 1. Google Prediction API: a Machine Learning black box for developers

- 내부 알고리즘을 알 수 없음
  - 알고리즘의 Parameter 설정을 할 수 없음
- => Data를 이용한 모델(블랙박스) 생성과 예측만이 가능함

<https://www.quora.com/What-algorithms-underlie-the-Google-Prediction-API>

### What algorithms underlie the Google Prediction

API?

Prediction API 알고리즘이 뭐니?

It says it "automatically selects from several available machine learning techniques", but does anyone have a good sense of what these techniques are?



Brandon Ballinger, applied ML to fraud, speech recognition, and ads

Written Jul 16, 2011

아마 Linear model인 것 같아..

Users have reported that the Prediction API has trouble learning XOR, which suggests a linear model--I'd guess they've implemented the most common linear classifiers like logistic regression, naive bayes, and SVMs.



Zachary Goldberg

Written Oct 13, 2011

우리는 여러 알고리즘을 가지고 있고 추가해가고 있어. 어려움이 있으면 연락 줘.

Hi! PM for the Prediction API at Google here :). We have quite a few algorithms underlying the Prediction API and we're constantly adding new ones. If you have a use case that you're having trouble getting good results with please do write to us at [prediction-api-discuss@googlegro...](mailto:prediction-api-discuss@googlegro...) and we'll see if we can find out why!

Happy Predicting,

-Zach

## 2. Less friendly Interface

- RESTful API

- Not Browser Interface

=> 그래도 Explorer API 때문에 단건 데이터에 대해 요청을 보낼 때는 비교적 쉽고, JSON 파일을 해석하는 것이 어렵지 않고 파싱(parsing)도 하면 됨

Search Result > Prediction API v1.6 > prediction.trainedmodels.predict

Authorize requests using OAuth 2.0:  OFF 

**project**

The project associated with the model. (string)

**id**

The unique name for the predictive model. (string)

비교적 간단!!

fields

Selector specifying which fields to include in a partial response.  
[Use fields editor](#)

Request body

```
{
}
```

**bold red** = required

Authorize and execute

[Execute without OAuth](#)

200

- Hide headers -

```
cache-control: private, max-age=0, must-revalidate, no-transform
content-encoding: gzip
content-length: 198
content-type: application/json; charset=UTF-8
date: Tue, 14 Mar 2017 05:59:36 GMT
etag: "LSfY9w58cvTpeojxLkvYtx4_xj4/qz17r0m26GPR0scUf73mEYLulNc"
expires: Tue, 14 Mar 2017 05:59:36 GMT
server: GSE
vary: Origin, X-Origin
```

```
-{
  "kind": "prediction#training",
  "id": "random-regression",
  "selfLink": "https://www.googleapis.com/prediction/v1.6/projects/test-project-160901/trainedmodels/random-regression",
  "created": "2017-03-13T23:36:19.847Z",
  "trainingStatus": "RUNNING"
}
```

## 3. Language support

- 어떤 언어이든 사용 가능? R이나 Matlab 등등 mathwork이나 분석을 위한 언어는 API 쓰기 어려울 것으로 판단됨

- sample code가 있는 언어는 Go, Java, .NET, Node.js, PHP, Python, Ruby

=> 웹을 다루기(?) 위해서 Python 역량을 키울 필요가 있다고 판단됨

## Start by programming language

You can use your favorite programming language on Cloud Platform, including the following languages.

|             |   |             |  |               |  |
|-------------|---|-------------|--|---------------|--|
| <b>GO</b>   | Go<br><a href="#">VIEW GO DOCS</a>        | <b>JAVA</b> | Java<br><a href="#">VIEW JAVA DOCS</a> | <b>.NET</b>   | .NET<br><a href="#">VIEW .NET DOCS</a>     |
| <b>NODE</b> | Node.js<br><a href="#">VIEW NODE DOCS</a> | <b>PHP</b>  | PHP<br><a href="#">VIEW PHP DOCS</a>   | <b>Python</b> | Python<br><a href="#">VIEW PYTHON DOCS</a> |
| <b>RUBY</b> | Ruby<br><a href="#">VIEW RUBY DOCS</a>    |             |  |               |  |

## 4. Not for analysis, For App

- 내부 알고리즘이 비공개인 것으로 판단되고, 파라미터(parameter) 조정도 할 수 없기 때문에 분석이나 실험을 위해 쓰기는 어려움

- 분석에 이용한다면, missing value 처리에 쓸 수도 있으나 모델이 비공개이므로 공격의 가능성 있음

- 빅데이터 분석 platform, service 또는 App을 위한 API임

=> 빅데이터 분석 platform, service 또는 App을 위해서 이용되기 좋음

[https://cloud.google.com/prediction/docs/smart\\_autofill\\_add\\_on](https://cloud.google.com/prediction/docs/smart_autofill_add_on)

The screenshot displays the Google Sheets interface with the Smart Autofill add-on. The add-on menu is open, showing options like 'Start', 'Help', 'Get add-ons...', and 'Manage add-ons...'. The main spreadsheet contains a table with columns: Price, Year, Num. of Miles, Num. of Doors, and Type. The 'Price' column has a value of 3363.892771 highlighted in blue. The 'Type' column lists various vehicle types like Car, Truck, and 2-Door. On the right side, there is a 'Smart Autofill' sidebar with 'Autofill Info' showing 'Average error: 2740.58', 'Number of pre-labeled rows: 17', and 'Number of rows filled: 5'. A 'Start New Autofill' button is visible at the bottom of the sidebar.

| Price    | Year | Num. of Miles | Num. of Doors | Type  |
|----------|------|---------------|---------------|-------|
| 22000    | 2012 | 13000         | 2-Door        | Car   |
| 14000    | 2010 | 30000         | 2-Door        | Car   |
| 12000    | 2010 | 60000         | 2-Door        | Car   |
| 13000    | 2010 | 73500         | 4-Door        | Car   |
| 10467.12 | 2009 | 70000         | 4-Door        | Truck |
| 9500     | 2009 | 78000         | 4-Door        | Car   |
| 9000     | 2009 | 78000         | 4-Door        | Car   |
| 14229.55 | 2006 | 47000         | 2-Door        | Car   |
| 4000     | 2006 | 124000        | 2-Door        | Car   |
| 6000     | 2005 | 28800         | 4-Door        | Car   |
| 6892.88  | 2005 | 131000        | 2-Door        | Truck |
| 6041.08  | 2004 | 119000        | 4-Door        | Truck |
| 3000     | 2004 | 131000        | 2-Door        | Truck |
| 2000     | 2004 | 119000        | 4-Door        | Truck |
| 3000     | 2003 | 177000        | 4-Door        | Car   |
| 2000     | 2004 | 209000        | 4-Door        | Truck |
| 3000     | 2003 | 130000        | 4-Door        | Truck |
| 3000     | 2003 | 138000        | 2-Door        | Car   |
| 1900     | 2003 | 160000        | 4-Door        | Car   |
| 2500     | 2003 | 190000        | 2-Door        | Truck |
| 5000     | 2002 | 135000        | 4-Door        | Car   |
| 1800     | 2001 | 62000         | 4-Door        | Car   |
| 1300     | 1997 | 163000        | 2-Door        | Truck |
| 13000    | 1997 | 138000        | 4-Door        | Car   |

## 5. 다른 API

- API마다 parameter의 차이는 있을 수 있지만, 다른 API들도 작동원리, 인증 방식 등이 같은 것으로 판단됨
  - 데이터가 있고, 분석하고 싶은 API가 있을 경우 다른 API도 쉽게 이용 가능할 것으로 판단됨
- => 다른 API들도 쉽게 이용할 수 있을 것으로 판단됨



## Machine Learning

[Cloud Machine Learning <sup>Beta</sup>](#)

[Vision API](#)

[Speech API <sup>Beta</sup>](#)

[Natural Language API](#)

[Prediction API](#)

[Translation API](#)

[Less](#)

# Pricing

- Usage fees:

- Prediction:

- 10,000 predictions/month: \$0.00
- 10,001+ predictions/month: \$0.50/1,000 predictions beyond the initial 10,000

- Training:

- \$0.002/MB bulk trained (maximum size of each dataset: 2.5GB)
- 0-10,000 streaming updates: \$0.00
- 10,001+ streaming updates: \$0.05/1,000 updates beyond the initial 10,000.

- Usage limits:

- If you intend to make more than 40,000 predictions/day, [please contact us](#).
- Google Prediction has a default limit of 2,000,000 predictions/day per project for your own models.
- Hosted models have a usage limit of 1 Storage models. Developers can set higher paid model's documentation for more details



| Multi-Regional Storage<br>(per GB per Month) | Regional Storage<br>(per GB per Month) | Nearline Storage<br>(per GB per Month) | Coldline Storage<br>(per GB per Month) |
|--|--|--|--|
| \$0.026                                      | \$0.02                                 | \$0.01                                 | \$0.007                                |

## Japan (asia-northeast1) Region Pricing

| Regional Storage<br>(per GB per Month) | Nearline Storage<br>(per GB per Month) | Coldline Storage<br>(per GB per Month) |
|--|--|--|
| \$0.023                                | \$0.016                                | \$0.01                                 |

Thank you.